Mini Project Report

on

EFFICIENT HUMAN VIOLENCE DETECTION

By

TUSAR AGARWAL (Reg. No.-202000219)

NILESH KUMAR (Reg. No.-202000459)

KESHAV KUMAR (Reg. No.-202000481)

Group Id – C17

In partial fulfillment of requirements for the award of degree in

Bachelor of Technology in Computer Science and Engineering

(2023)

Under the Project Guidance of

Mr. Chinmoy Kar

Assistant Professor - I

Sikkim Manipal Institute of Technology, Majitar

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent college of Sikkim Manipal University)

MAJITAR, RANGPO, EAST SIKKIM – 737136

PROJECT COMPLETION CERTIFICATE

This is to certify that the below mentioned students of Sikkim Manipal Institute of Technology have worked under my supervision and guidance from **9th January 2023 to 29th April 2023** and successfully completed the Mini project entitled **“EFFICIENT HUMAN VIOLENCE DETECTION”** in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

University Registration No	Name of Student	Course
202000219	**TUSAR AGARWAL**	**B.Tech (CSE)**
202000459	**NILESH KUMAR**	**B.Tech (CSE)**
202000481	**KESHAV KUMAR**	**B.Tech (CSE)**

Mr. Chinmoy Kar

Assistant Professor - I

Department of Computer Science and Engineering

Sikkim Manipal institute of Technology

Majhitar, Sikkim – 737136

PROJECT REVIEW CERTIFICATE

This is to certify that the work recorded in this project report entitled **"EFFICIENT HUMAN VIOLENCE DETECTION"** has been jointly carried out by **TUSAR AGARWAL(Reg. 202000219), NILESH KUMAR (Reg. 202000459) and KESHAV KUMAR (Reg. 202000481)** of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering. This report has been duly reviewed by the undersigned and recommended for final submission for Mini Project Viva Examination.

Mr. Chinmoy Kumar

Assistant Professor - I

Department of Computer Science and Engineering

Sikkim Manipal Institute of Technology

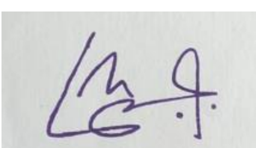
Majhitar, Sikkim – 737136

CERTIFICATE OF ACCEPTANCE

This is to certify that the below mentioned students of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology (SMIT) have worked under the supervision of **Mr. Chinmoy Kar**, Assistant Professor - I, Department of Computer Science and Engineering from **9th January 2023 to 29th April 2023** on the project entitled **“EFFICIENT HUMAN VIOLENCE DETECTION”.**

The project is hereby accepted by the Department of Computer Science & Engineering, SMIT in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

University Registration No	Name of Student	Project Venue
202000219	**Tusar Agarwal**	
202000459	**Nilesh Kumar**	**SMIT**
202000481	**Keshav Kumar**	



Dr. Udit Kumar Chakraborty

Professor & Head of the Department

Computer Science & Engineering Department

Sikkim Manipal Institute of Technology

Majhitar, Sikkim – 737136

DECLARATION

We, the undersigned, hereby declare that the work recorded in this project report entitled "**EFFICIENT HUMAN VIOLENCE DETECTION**" in partial fulfillment for the requirements of award of B.Tech (CSE) from Sikkim Manipal Institute of Technology (A constituent college of Sikkim Manipal University) is a faithful and bonafide project work carried out at "**SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY**" under the supervision and guidance of **Mr. Chinmoy Kar,** Assistant Professor, Department of Computer Science and Engineering.

The results of this investigation reported in this project have so far not been reported for any other Degree or any other technical forum.

The assistance and help received during the course of the investigation have been duly acknowledged.

TUSAR AGARWAL (Reg. No.-202000219)

NILESH KUMAR (Reg. No.-202000459)

KESHAV KUMAR (Reg. No.-202000481)

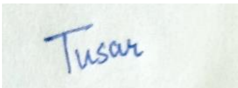
ACKNOWLEDGMENT

We take this opportunity to acknowledge indebtedness and a deep sense of gratitude to our guide **Mr. Chinmoy Kar** for his valuable guidance and supervision throughout the project work which shaped the present work as it shows.

We pay our deep sense of gratitude to **Prof. (Dr.) Udit Kumar Chakraborty, HOD, Computer Science & Engineering Department, Sikkim Manipal Institute of Technology** for giving us the opportunity to work on this project and providing all support required.

We are obliged to our project coordinators **Dr. Sandeep Gurung** and **Mr. Biraj Upadhyaya** for elevating, inspiration and supervising in completion of our project.

We would also like to thank any other staff of **Computer Science & Engineering Department, Sikkim Manipal Institute of Technology** for giving us continuous support and guidance that has helped us in completion of our project.

TUSAR AGARWAL (Reg. No.-202000219) 

NILESH KUMAR (Reg. No.-202000459) 

KESHAV KUMAR (Reg. No.-202000481) 

DOCUMENT CONTROL SHEET

1	Report No	CSE/Mini Project/Internal/B.Tech/C/Group 17/2023
2	Title of the Report	EFFICIENT HUMAN VIOLENCE DETECTION
3	Type of Report	Technical
4	Author	TUSAR AGARWAL, NILESH KUMAR, KESHAV KUMAR
5	Organizing Unit	Sikkim Manipal Institute of Technology
6	Language of the Document	English
7	Abstract	This project is aimed to present a violence detection system, aimed at detecting instances of aggressive human behaviour in public environment. This report introduces a method to detect violence in humans using deep learning. The method uses a transfer learning model to detect frames that contain violent activity from a video and classify those frames.
8	Security Classification	General
9	Distribution Statement	General

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The project “EFFICIENT HUMAN VIOLENCE DETECTION ” aims to develop a convolutional neural network (CNN) model using transfer learning to detect and classify violent and non-violent frames from the input video. The dataset used in this project consists of videos from two different classes: violent and non-violent. The videos from the dataset are converted to frames and preprocessed before training the model. The model is optimized using the adam optimizer and binary cross-entropy loss function. Key frames are extracted from the video stream using methods such as optical flow and color histogram analysis, which reduces the computational overhead and enhances the effectiveness of the model. The model is trained using transfer learning, where the pre-trained MobileNet model is fine-tuned on the violence detection task. The results of the project show that the MobileNet model can achieve high accuracy in classifying violence, with less computational time and cost when combined with a key frame extraction algorithm. The implementation of this model has the potential to improve public safety and reduce the spread of harmful content on the internet.

1. INTRODUCTION

The EFFICIENT HUMAN VIOLENCE DETECTION project is an application of transfer learning technique for violence detection . The rate of crimes is increasing day by day in all societies in the world. And the governments are doing everything they could do, spending huge amount of budget to deploy safety surveillance systems & man-power to monitor them & identify the criminals. That's a great thing to do, but the way it is done, can be more efficient.

Many crimes are committed every day, and probably hundreds are occurring right now in the world. Automating this task using Human Violence Detection System will not only save time but also reduce man-power thus reducing the effective cost.

Convolutional Neural Networks (CNN) are a type of deep learning model that has shown remarkable success in detecting and classifying images. CNN models have been applied to a wide range of tasks, including image recognition, object detection, and facial recognition. In recent years, there has been increasing interest in using CNN models for violence detection. By analyzing video feeds and identifying patterns associated with violent behavior, these models can provide valuable insights into potential incidents and alert authorities in real-time. This report will examine the application of CNN models using transfer learning to detect human violence.

2.LITERATURE SURVEY

Table 1

SI No	Author	Paper and publication Details	Findings	Relevance to the project
1.	Muhammad Ramzan, Adnan Abid, Hikmat UllahKhan, Shahid Mahmood Awan	"A Review on State-of-the-Art Violence Detection Techniques," in *IEEE Access*, vol. 7, pp. 107560-107575, 2019, doi: 10.1109/ACCESS.2019.2932114	The use of machine learning and deep learning techniques have shown promising results for violence detection, with high accuracy and low false positive rates.	Help us to provide a broader context for the project, and help to identify potential opportunities for future research in this field.
2.	A. Mumtaz, A. B. Sargano and Z. Habib	"Violence Detection in Surveillance Videos with Deep Network Using Transfer Learning," *2018 2nd European Conference on Electrical Engineering and Computer Science (EECS)*, Bern,	Used transfer learning to fine-tune a pre-trained convolutional neural network (CNN) to classify frames of the video as violent or non-violent.	Utilizing pre-trained convolutional neural network models such as VGG-16 and Inception-v3, the proposed approach can effectively learn and classify different features related to violence in the surveillance videos.

3.	Khan, S.U.; Haq, I.U.; Rho, S.; Baik, S.W.; Lee, M.Y.	Cover the Violence: A Novel Deep-Learning-Based Approach Towards Violence-Detection in Movies. *Appl. Sci.* **2019**, *9*, 4963. https://doi.org/10.3390/app9224963	The proposed approach uses a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to capture both spatial and temporal features of violent scenes in movies	Although the proposed approach does not explicitly use transfer learning, it is still relevant to a transfer learning-based project as it can provide valuable insights and techniques for improving the performance of the model.
4.	M. I. H. Azhar, F. H. K. Zaman, N. M. Tahir and H. Hashim,	"People Tracking System Using DeepSORT," 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2020, pp. 137-141, doi: 10.1109/ICCSCE50387.2020.9204956.	DeepSORT has been shown to perform well on a variety of dataset and is computationally efficient and can run in real-time on standard hardware	The use of deep learning techniques can improve the accuracy of person reidentification which can result in high accuracy of people detection and tracking.

`

3.PROBLEM DEFINITION

- Traditional systems rely on human operators to monitor multiple video streams of specific events.
- Due to the large number of video streams and limited human concentration abilities, the chance of an incident actually being noticed may be much lower than one might expect.
- In a complex environment containing multiple individuals present various activities may be occurring. A way is needed to distinguish the violent activities from the normal, everyday activities.
- The variation in body movements and image quality in the input frames can make the classification challenging.

4.SOLUTION STRATEGY

1. **Data Collection:** Collecting a dataset of video clips containing violent and non-violent scenes. The dataset contains 2000 violent and non-violent videos.

2. **Data Preprocessing:** Converting the videos to frames and extracting features from the frames. Resizing the frames to a fixed size to fit the MobileNet architecture. Splitting the dataset into training, validation, and test sets.

3. **Using appropriate optimization techniques:** Using optimization techniques such as the adam optimizer and the binary cross-entropy loss function can help improve the accuracy of the model.

4. **Model Selection:** MobileNet architecture is chosen as the base model and added a few layers on it for classification. The MobileNet architecture is suitable for the project as it is lightweight and has low computational cost.

5. **Model Training:** Training the model using the training set and validate it using the validation set.

5.DESIGN

Semantic diagram

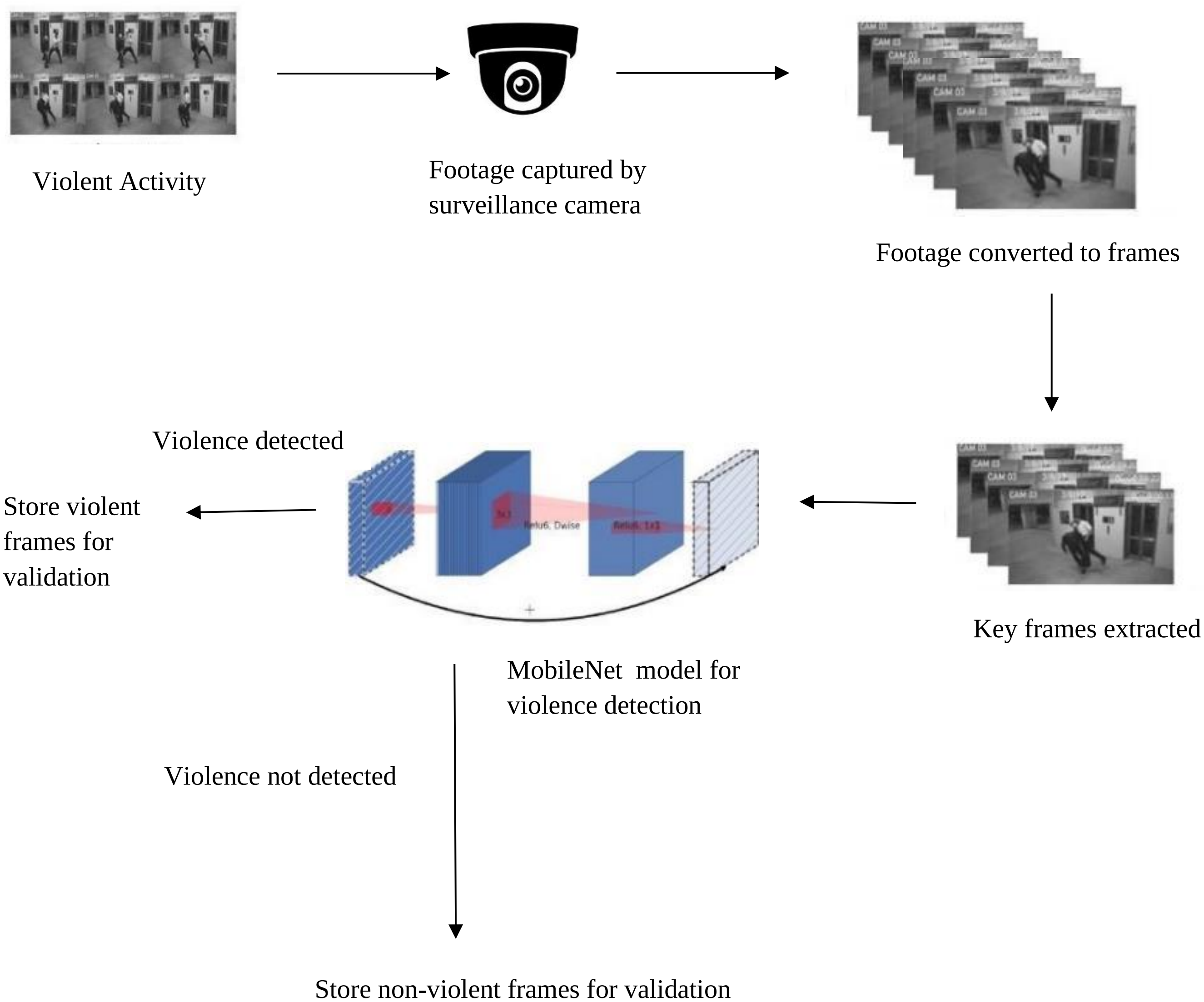


Fig.(1.1)

Block diagram

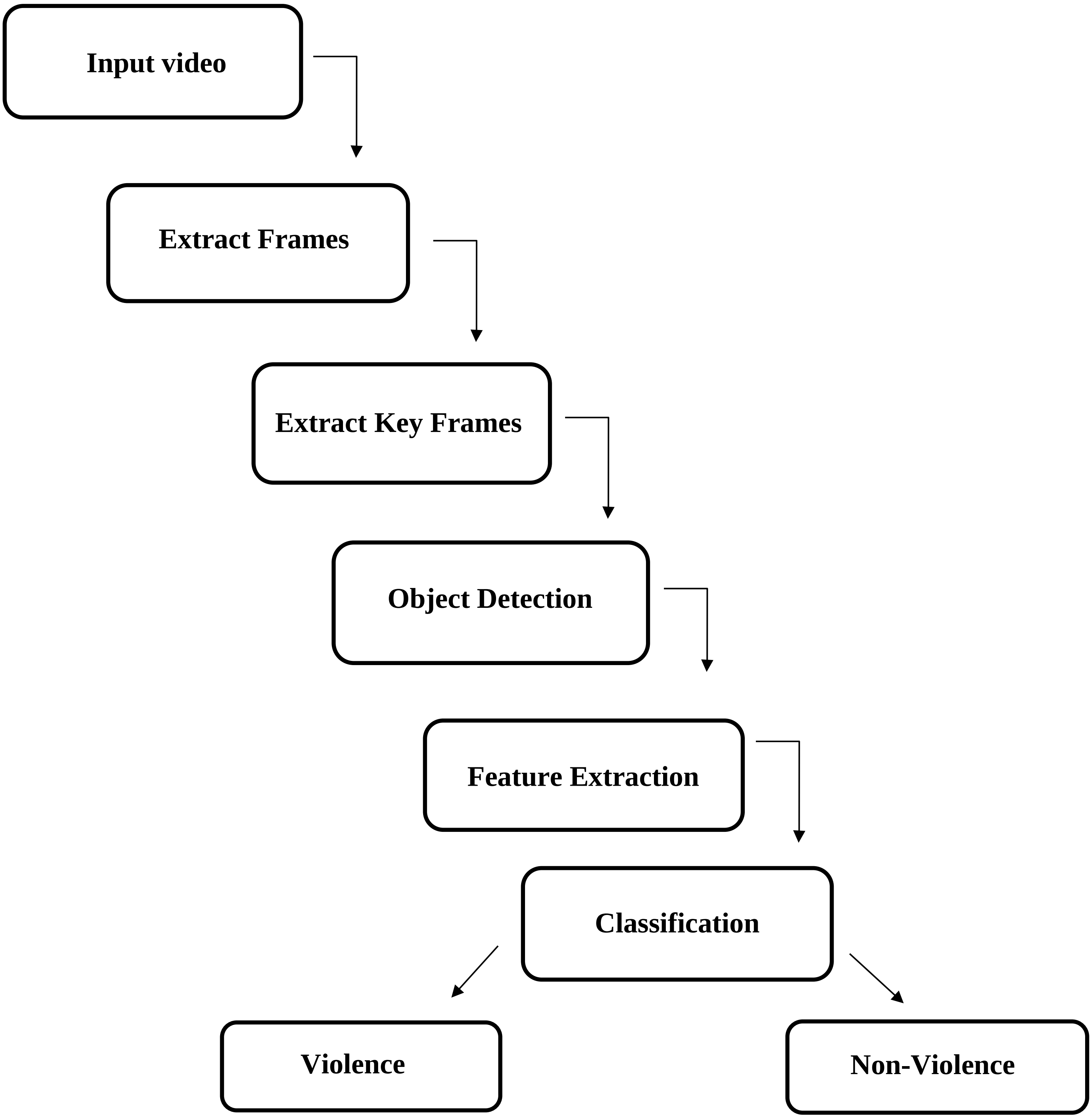


Fig.(1.2)

6.IMPLEMENTATION DETAILS

- **Transfer learning using MobileNet architecture** :

Transfer learning is a technique in machine learning where a pre-trained model is used as a starting point for a new model that is trained on a different but related task or dataset. The idea behind transfer learning is that the knowledge learned by a model on one task can be leveraged to improve the performance of the model on another task.

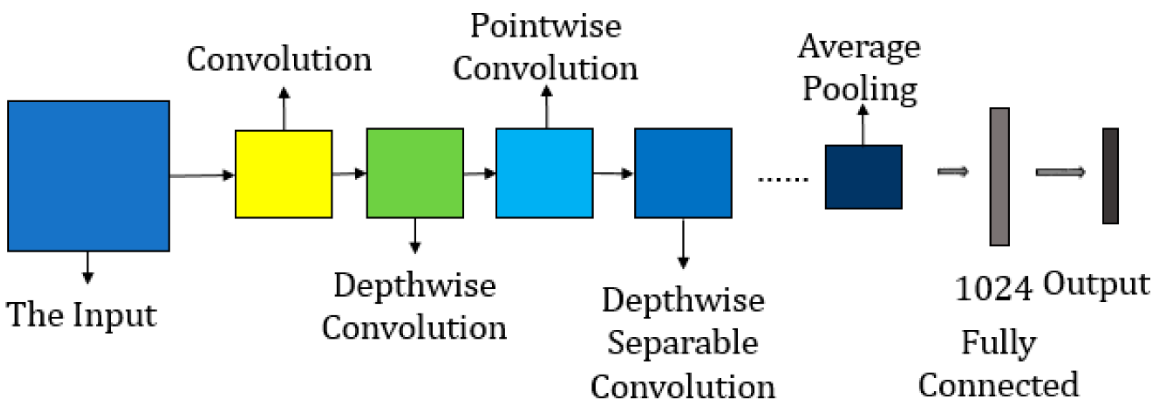


Fig.(1.3) MobileNet architecture

MobileNet architecture:

MobileNet is a convolutional neural network architecture designed to efficiently process images on mobile devices with limited computational resources.

MobileNet architecture consists of several layers that are used to process input images and produce the final output. Here is a brief overview of the layers in MobileNet:

1. **Input layer:** This is the layer that takes the input image and passes it to the first convolutional layer.

2. **Convolutional layers:** The MobileNet architecture consists of several convolutional layers that use depthwise separable convolutions to extract features from the input image. These layers are designed to be computationally efficient and have a low number of parameters.

3. **Batch normalization:** This layer is used to normalize the output of the convolutional layers, which helps improve the accuracy and stability of the network.

4. **ReLU activation:** This layer applies the Rectified Linear Unit activation function to the output of the batch normalization layer. This non-linear function introduces non-linearity into the network and helps to make the output more discriminative.

5. **Depthwise convolutional layer:** This layer performs spatial filtering of the input image by applying a depthwise convolution operation.

6. **Pointwise convolutional layer:** This layer performs cross-channel filtering of the input image by applying a pointwise convolution operation.

7. **Global average pooling:** This layer takes the output of the last convolutional layer and computes the average value of each feature map. This reduces the dimensionality of the output and produces a feature vector that is used as input to the final fully connected layer.

Dataset

The name of the dataset that is used in this project is "**Real life violent situations dataset**"(Source – https://www.kaggle.com/code/chaitanya99/real-life-violence-detection-mobilenetv2/input). This dataset includes 2000 video clips that are categorized into two classes: violence and non-violence. The videos have an average duration of 5 seconds and are primarily obtained from CCTV footages. During the training process, we selected 1000 video clips from both the violent and non-violent classes at each epoch to ensure balanced representation and to avoid any biases towards either class.

The videos from the dataset were first converted into frames. The total number of frames obtained was 87,568. To train the model, we fed 80% of these frames into the MobileNet model. We trained the model for 10 epochs and achieved an accuracy of 96%.

Key frame extraction

We implemented a two-step approach to classify videos as violent or non-violent. First, we used the key frame algorithm to extract key frames from the input video. These key frames were then fed into the MobileNet model for classification.

To evaluate the effectiveness of our approach, we compared it with another approach that did not use key frame extraction. The results showed that our approach had higher accuracy and faster processing time compared to the other approach. This is because the key frame algorithm reduces the amount of data that needs to be processed, allowing for more efficient and accurate classification.

Overall, our study demonstrates the effectiveness of using key frame extraction and the MobileNet model for video classification tasks, particularly in scenarios where computational resources are limited. The findings have important implications for the development of intelligent video surveillance systems and other applications that require real-time video analysis.

Algorithm for key frame extraction

1. Load a pre-trained violence detection model.

2. Set the key frame interval to a certain value, which in this case is 10.

3. Set up output directories for violent and non-violent frames.

4. Open the input video file using OpenCV.

5. Initialize a frame counter to keep track of the number of frames processed.

6. Set the previous key frame to None.

7. While the video is still open, read the next frame.

8. If there are no more frames to read, break out of the loop.

9. Increment the frame counter.

10. If this is the first frame, set the previous key frame and current key frame to this frame.

11. If this is a key frame (i.e., frame_count is a multiple of key_frame_interval), calculate the mean absolute error (MAE) between the current frame and the previous key frame using cv2.absdiff().

12. If the MAE is greater than 0.05, set the previous key frame and current key frame to this frame.

13. Otherwise, continue to the next frame.

14. Classify the current key frame as violent or non-violent using the pre-trained model.

15. Save the current key frame in the appropriate output directory based on its classification.

16. Release the input video file and close all windows.

17. Print out statistics, including the total number of frames processed, the number of violent frames saved, and the number of non-violent frames saved.

TOOLS:

The prerequisites for this project are – Python, Google Colab and High-Speed Internet.

LIBRARIES:

- NumPy
- Tensorflow
- Keras
- OpenCV
- OS

Algorithm

1. Import required libraries
2. Define paths to the training set, test set, and prediction video
3. Define the data augmentation pipeline for the training set
4. Define the data generator for the test set
5. Define the MobileNet model with pre-trained weights
6. Add a few layers on top of MobileNet to create a new model
7. Train the model on the training set
8. Load the saved model for prediction
9. Define Twilio credentials and phone numbers
10. Define a function to classify frames as violent or non-violent using the mod

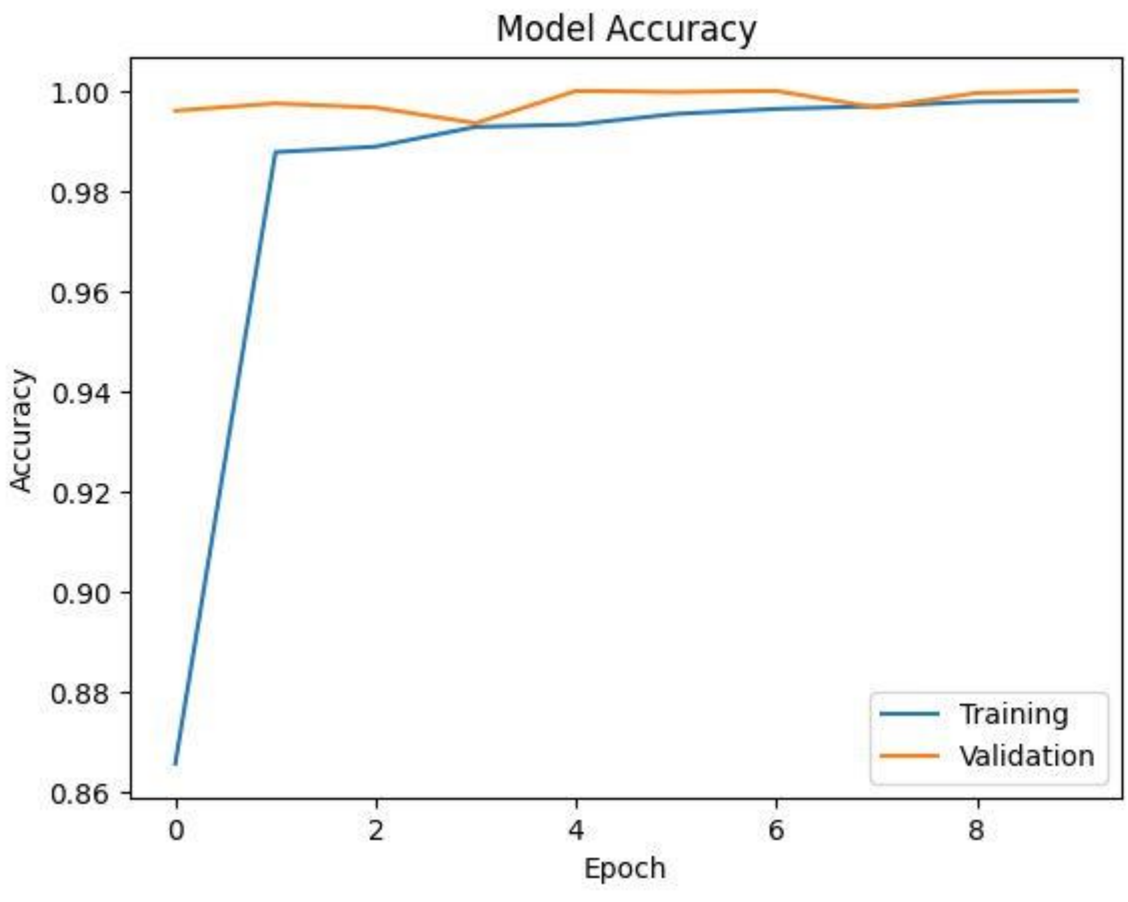
Epochs and Accuracy

Table 2

Epoch	Loss	Accuracy
1	0.2517	0.9621
2	0.1082	0.9896
3	0.0614	0.9928
4	0.0390	0.9949
5	0.0279	0.9956
6	0.0198	0.9967
7	0.0157	0.9969
8	0.0134	0.9973
9	0.0135	0.9970
10	0.0101	0.9977

7. RESULTS

GRAPHICAL REPRESENTATION:



Accuracy w.r.t Epoch

Analysis of Output from the model

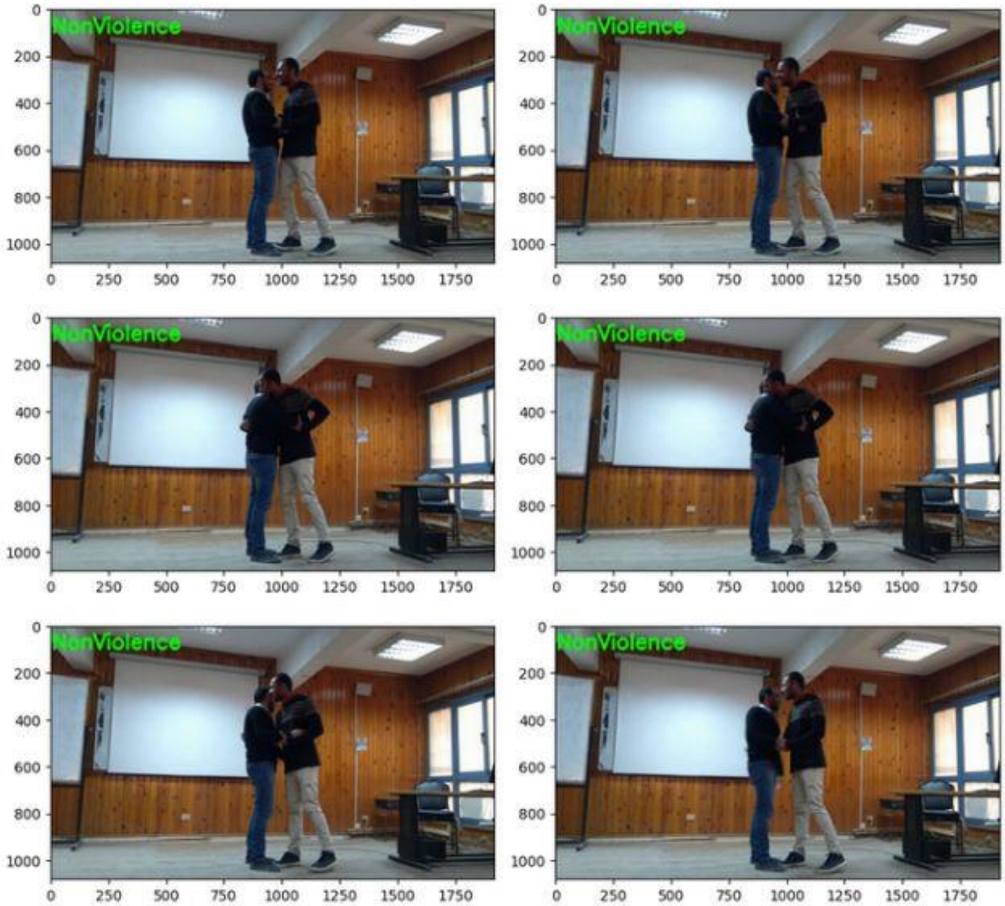
Table 3

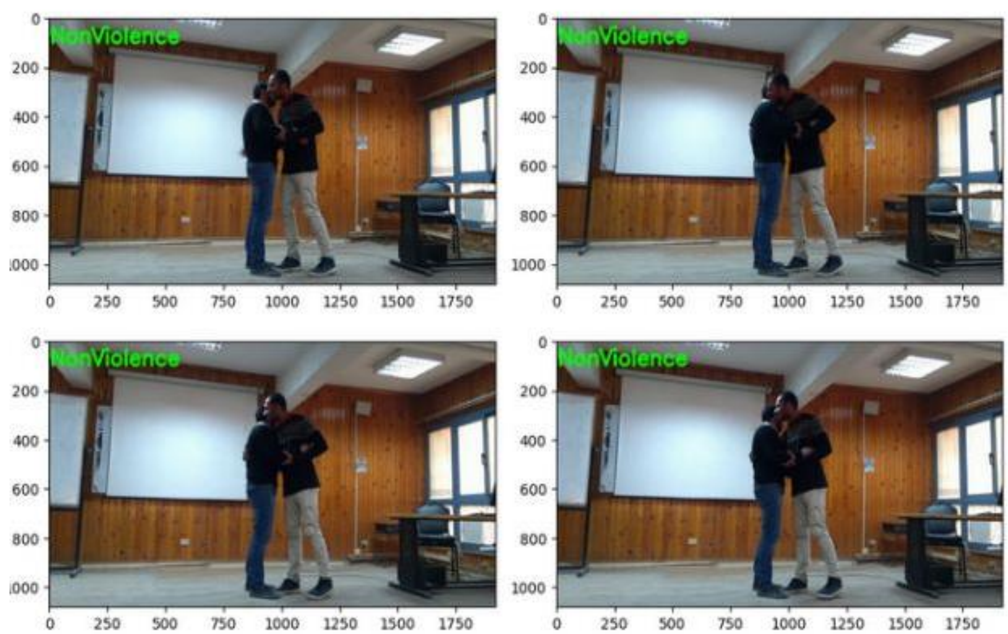
VIDEOS	Total frames processed	Violent frames	Non-violent frames	false positive	false negative
hospital.mp4	1238	46	78	40	76
Climpchamp.mp4	280	10	19	8	12
Clipchamp_1.mp4	271	13	15	8	15
Clipchamp_2.mp4	329	14	19	9	19
Clipchamp_3.mp4	257	10	16	6	14
NV_200	125	0	13	0	13
NV_201	125	0	13	0	13
NV_202	113	0	12	0	12
NV_203	125	0	13	0	13
NV_204	125	0	12	0	12
V_301	132	14	0	10	0
V_302	129	13	0	10	0
V_303	183	19	0	18	0
V_304	156	16	0	13	0
V_305	129	13	0	13	0

ANALYSIS ON RESULT

Violent frames:

Non-Violent Frames:





8.CONCLUSION

The use of MobileNet to classify frames is a promising approach to enhancing public safety and security. The model can effectively identify violent incidents in video footage and provide real-time alerts to security personnel or law enforcement agencies. However, the model has its limitations, including limited accuracy, scope, sensitivity to input quality, limited context, and scalability.

Despite these limitations, the model has several strengths, including its ability to provide timely alerts, scalability, and ease of integration with existing security systems. Furthermore, the use of MobileNet provides a powerful and efficient image classification model that can effectively classify frames as violent or non-violent.

In the future, improvements to the violence detection model can be made by focusing on improving accuracy, adding contextual information, enabling real-time video analysis, supporting multi-camera scenarios, and improving scalability and deployment. Overall, a violence detection model using MobileNet has the potential to improve public safety and security and can be a valuable tool in preventing and responding to violent incidents.

8.LIMITATION

i. **Limited accuracy:** While MobileNet is a powerful image classification model, it may not always accurately classify violent and non-violent frames. This can result in false positives or false negatives, which can lead to inaccurate detection of violent incidents.

ii. **Limited context:** The violence detection model only considers individual frames in isolation and does not take into account the context of the video. For example, a violent incident may be part of a larger altercation that is not captured in the individual frames.

iii. **Sensitivity to input quality:** The accuracy of the violence detection model may be affected by the quality of the input video. Poor lighting conditions or low resolution may make it more difficult for the model to accurately classify frames.

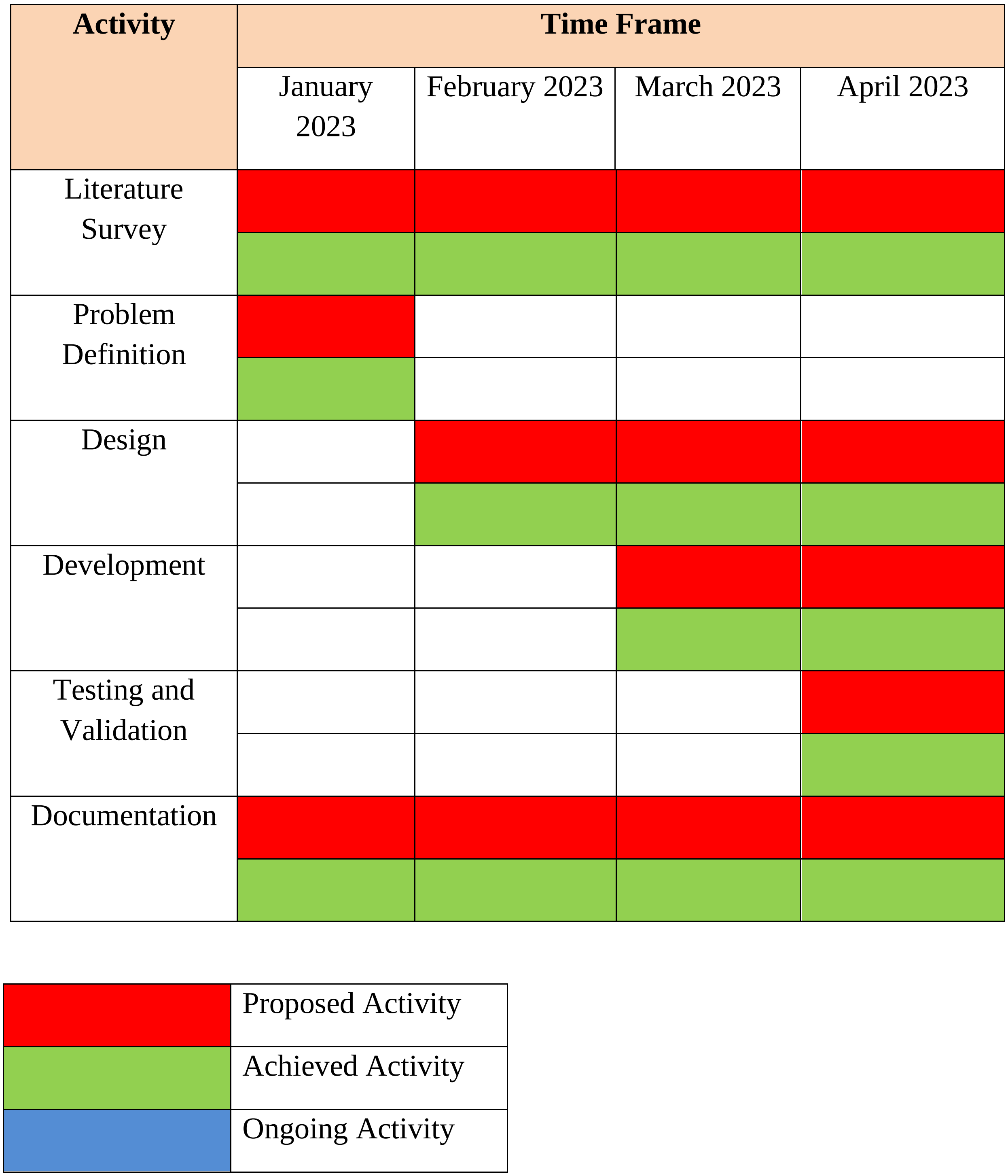
9.FUTURE SCOPE

1. **Context-aware violence detection:** Future work can explore the integration of contextual information in the violence detection model. This could involve analyzing the audio, metadata, or other information associated with the video to improve the accuracy of the model and better understand the context of the violence.

2. **Real-time video analysis:** The violence detection model can be improved by focusing on real-time video analysis. This could involve developing more efficient algorithms for video processing or leveraging hardware acceleration to improve the speed of the model.

3. **Scalability and deployment:** Future work can focus on developing a scalable and deployable violence detection model that can be integrated into existing security systems. This could involve developing an API or software development kit (SDK) for integration with existing surveillance systems, or developing a cloud-based platform for video analysis.

10.GANT CHART



References

[1] Muhammad Ramzan, Adnan Abid, et.. "A Review on State-of-the-Art Violence Detection Techniques.," *IEEE Access*, vol. 7,pp. 107560-107575,2019, no. doi:10.1109/ACCESS.2019.2932114, 2019.

[2] A. Mumtaz, A.B. Sargano and Z. Habib, "Violence Detection in Surveillance Videos with Deep Network Using Transfer Learning," *2018 2nd European Conference on Electrical and Computer Science(EECS)* ", Bern, 2018.

[3] Khan,S,U.; Haq,I.U.; Rho,S.; Baik, S.W.; Lee, M.Y., "Cover the Violence: A Novel Deep-Learning-Based Approach Towards Violence-Detection in Movies., "*Appl. Sci. 2019,9,4963.* " no. doi:10.3390/app9224963, 2019.

[4] M.I.H. Azhar, F.H.K. Zaman, N.M. Tahir and H. Hashim, "People Tracking System Using DeepSORT, "2020 10th IEEE International Conference on Control System,Computing and Engineering (ICCSCE), Penang, Malaysia,2020, pp. 137-141, no. doi:10.1109/ICCSCE50387.2020.9204956., 2020.

violence detection

ORIGINALITY REPORT

14%	12%	10%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Nanyang Technological University Student Paper	3%
2	Shaheena Noor, Maria Waqas, Muhammad Imran Saleem, Humera Noor Minhas. "Automatic Object Tracking and Segmentation using Unsupervised SiamMask", IEEE Access, 2021 Publication	2%
3	Submitted to University of Essex Student Paper	2%
4	www.mdpi.com Internet Source	2%
5	coek.info Internet Source	1%
6	Kim, Sang Hyun, Rae-Hong Park, Chung-Sheng Li, and Rainer W. Lienhart. "", Storage and Retrieval for Media Databases 2002, 2001. Publication	1%
7	iq.opengenus.org Internet Source	

		1%
8	us.grademiners.com Internet Source	1%
9	www.researchgate.net Internet Source	1%
10	github.com Internet Source	<1%
11	repository.ubn.ru.nl Internet Source	<1%
12	Muhammad Ramzan, Adnan Abid, Shahid Mahmood Awan. "Automatic Unusual Activities Recognition Using Deep Learning in Academia", Computers, Materials & Continua, 2022 Publication	<1%
13	mdpi-res.com Internet Source	<1%
14	Sarah Azimi, Corrado De Sio, Francesco Carlucci, Luca Sterpone. "Fighting for a future free from violence: A framework for real-time detection of “Signal for Help”", Intelligent Systems with Applications, 2023 Publication	<1%