Real Time Violence Detection: CNN

Veena Sharma
Department of CSE(Data Science),
A.P. Shah Institute of Technology,
Thane (M.H), India 400615
Email:veenasharma444@apsit edu.in

Pooja Kumbhar
Department of CSE(Data Science),
A.P. Shah Institute of Technology,
Thane (M.H), India 400615
Email:poojakumbhar416@apsit.edu.in

Sanika Shelke
Department of CSE(Data Science),
A.P. Shah Institute of Technology,
Thane (M.H), India 400615
Emailsanikashelke461@apsit.edu.in

Prof,Sarala Mary
Department of CSE(Data Science),
A.P. Shah Institute of Technology,
Thane (M.H), India 400615
Email: saralamary@gmail.com

Abstract— This paper presents a violence detection system integrated into a web-based platform for real-time analysis of uploaded images. Leveraging advanced image processing techniques and machine learning algorithms, the system autonomously evaluates the content of uploaded images to determine the presence or absence of violent behavior. Upon uploading an image, the system rapidly extracts relevant visual features and employs a trained classification model to classify the image as either depicting violence or not. With the rise in violent incidents in public spaces underscores the critical need for effective real-time violence detection systems. This research paper presents a novel approach leveraging Convolutional Neural Networks (CNNs) networks for the real-time detection of violent activities in surveillance videos. The proposed system integrates the spatial and temporal features of the video frames to accurately identify violent behaviours. Initially, the CNN extracts spatial features from individual frames, capturing relevant visual patterns indicative of violence. The model is trained on a large-scale dataset of annotated surveillance videos, facilitating robust learning of violence-related features. Experimental results demonstrate the efficacy of the proposed approach in accurately detecting violent activities in real-time scenarios, achieving superior performance compared to existing methods. The proposed system offers promising implications for enhancing public safety and security in various environments, including transportation hubs, public gatherings, and urban areas.

Keywords -Convolutional Neural Networks (CNN), Video analysis, Neural Network, Image analysis

I. INTRODUCTION

Real-time violence detection is a cutting-edge software program that leverages deep learning technology to proactively identify and respond to instances of violence in real-world environments. In an era where ensuring public safety is paramount, this innovative system represents a significant advancement in surveillance and security measures. By harnessing the power of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, the program continuously analyze live surveillance video feeds to detect and classify various forms of violent behaviour as they unfold. This proactive approach enables security personnel and law enforcement authorities to swiftly intervene and mitigate potential threats, thereby enhancing overall public safety.

The system operates by processing video streams in real-time, extracting spatial features from individual frames using ‘ CNNs, and capturing temporal dependencies through LSTM networks. This dual approach allows for the accurate identification of dynamic patterns associated with violent activities, including assaults, fights, and vandalism. By integrating advanced deep learning methodologies with annotated datasets and refined training techniques, theprogram continually enhances its detection capabilities and adapts to diverse environments and scenarios.

Overall, the real-time violence detection system represents a crucial technological innovation in the realm of public safety, offering a proactive solution to address the ever-evolving challenges posed by violence in modern society.. To continue with, this research paper further focuses on the related work, system architecture, analysis on the results achieved and future scope.

II. LITERATURE REVIEW

The literature on real-time violence detection in images and machine learning techniques provides valuable insights into the advancements, challenges, and potential applications of this technology. Several studies have explored various approaches and methodologies for automating the process of detecting violent behaviors in surveillance footage. Cruz et al. (2020) proposed a CNN-based approach for real-time detection of violent altercations in public spaces [2].Their study focused on developing a robust model capable of accurately identifying instances of physical aggression, such as fights and assaults, in crowded environments. By leveraging deep learning techniques and largescale annotated datasets, the authors achieved high accuracy and real-time performance, demonstrating the effectiveness of CNNs in addressing complex challenges. Silva Deena J, MD. Tabil Ahammed, Udaya Mouni Boppana, Maharin Afroj, Sudipto Ghosh, Sohaima Hossain[2022]presented an video representation learning for cctv based violence detection[3]. Ability to automatically recognize violence behaviors is one of the key technology for CCTV cameras. However, it is still a challenging task to obtain effective features for detecting violence in CCTV videos due to the visual quality of the video data. we propose a novel representation learning approach to improve the detection rate of violent behaviors in videos. Our 11 proposed approach consists of two parts. In the first part, we leverage features extracted from image-based deep convolution neural network to describe spatial information in a video frame. Nandini Bagga; Gajan Singh; Balamurugan Balusamy; Ajay Shanker Singh[2023] presented in Violence Detection in Real Life Videos using Convolutional Neural Network[4]. This research

paper explores how violence must be identified on real-time films taken by numerous surveillance cameras at all times and in all locations, which makes it difficult to perform. The methodology involves For image categorization, there are numerous pre-trained convolutional neural networks available. The result showcases a comprehensive understanding of violence detection and their pivotal role in information retrieval. Waseem Ullah,Amin Ullah,Ijaz Ul Haq,Khan Muhammad,Muhammad Sajjad,Sung Wook Baik[2021] presented CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks[5]. In this paper, we present an efficient deep features-based intelligent anomaly detection framework that can operate in surveillance networks with reduced time complexity. In current technological era, surveillance systems generate an enormous volume of video data on a daily basis, making its analysis a difficult task for computer vision experts. Manually searching for unusual events in these massive video streams is a challenging task, since they occur inconsistently and with low probability in real-world surveillance. The results of We performed extensive experiments on various anomaly detection benchmark datasets to validate the functionality of the proposed framework within complex surveillance scenarios.

III. PROPOSED SYSTEM

The aim of our project is to develop a real-time violence detection system that enhances public safety and security through proactive threat identification and intervention. Leveraging advanced algorithm libraries such as OpenCV for webcam integration, Plyer and Pygame for notification and alarm sound, and Convolutional Neural Networks (CNNs) for violence detection, our system aims to analyze live surveillance footage in real-time to identify instances of violence and trigger timely alerts. The system's core component include CNN Model, Image Preprocessing Algorithm, Plyer Module, Pygame Module, OpenCV Module, User Interface (UI), Data storage and Management.
Convolutional Neural Networks (CNNs) are a class of deep neural networks tailored for processing visual data. These networks comprise interconnected layers, including convolutional, pooling, and fully connected layers. In CNNs, convolutional layers apply filters to input data, extracting features that capture spatial patterns. Pooling layers then downsample the feature maps, reducing computational complexity while preserving essential information. CNNs excel in tasks such as image recognition, object detection, and classification, owing to their ability to capture spatial relationships in data

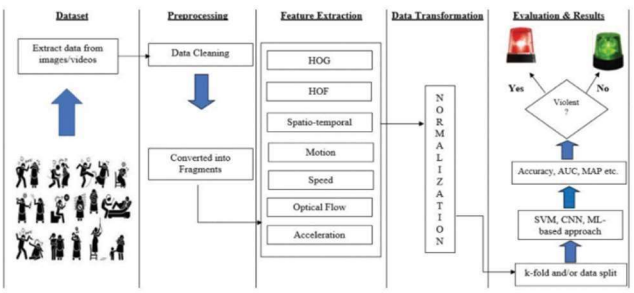


Figure 1: Architectural representation of CNN algorithm

Key components of the proposed system include:Web-based Interface: The system will feature an intuitive web interface accessible to users for uploading images and accessing violence detection functionalities. The interface will be designed for ease of use, allowing users to submit images effortlessly and receive prompt analysis results.Image Processing Pipeline: Upon image submission, the system will employ a sophisticated image processing pipeline to extract relevant visual features and patterns indicative of violent behavior.Techniques such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Convolutional Neural Networks (CNNs) will be utilized to capture both low-level and high-level features from the images. Machine Learning Models: Extracted features will serve as input to machine learning models trained to classify images as either depicting violence or not. Supervised learning algorithms such as Support Vector Machines (SVM), Random Forests, or deep learning architectures like CNNs will be employed to learn discriminative patterns from annotated datasets of Violent and non-violent images.

IV. RESULT AND ANALYSIS

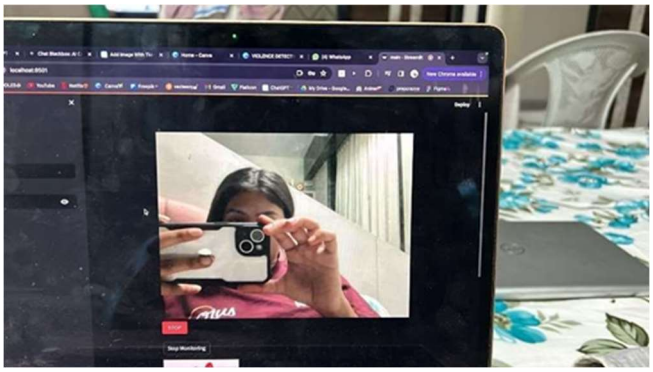
The home page of our application serves as the primary gateway, offering users the option to log in as administrators. Upon selecting the "Admin Login" button, administrators are prompted to authenticate their identity through a username and password. Upon successful authentication, administrators are directed to the admin dashboard, a centralized control panel housing various functionalities for managing the application.

FIGURE 2
Home Page

Among these features, administrators are presented with a pivotal control: the ability to manipulate the webcam functionality. Within the admin dashboard, two prominent buttons are displayed: "Start" and "Stop." These buttons enable administrators to initiate or terminate the webcam feed, respectively. When the "Start" button is activated, the application accesses the webcam hardware, initiating a live stream of the camera feed.

Figure 3 :Image uploa

This real-time feed can be viewed either within a designated area on the admin dashboard or in a separate window, allowing administrators to monitor events or activities in thecamera's vicinity. Conversely, clicking the "Stop" button halts the webcam functionality, terminating the live stream. This straightforward interface empowers administrators with the flexibility to activate or deactivate the webcam as needed, facilitating efficient surveillance or monitoring tasks. It's imperative to integrate robust security measures toensure that



only authorized administrators can access the admin dashboard and control the webcam functionality, safeguarding the integrity and privacy of the application.

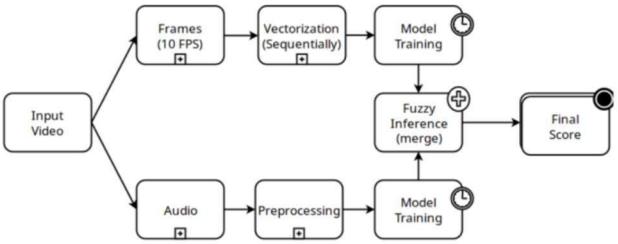
FIGURE 4 :Webcam Page

V.FUTURE SCOPE

Drawing inference from Audio: I can have an inference engine running, which uses some fuzzy logic that uses the result vector from both, video inference and audio inference, and using some composition (like min-max) draw a better inference. Now there are two approaches that I can think of that can help in classifying audio between two classes, violent and non-violent:

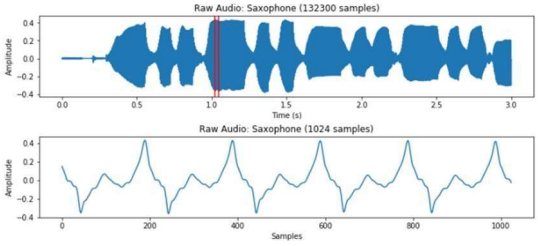
- Raw wave input to CNN 1DConv
- Mel Spectrogram transform input to CNN 2DConv

FIGURE 5
INFERENCE MODEL



Apparently, Mel Spectrogram approach has been more accurate as tested on some standard audio dataset such as Urban Sound Tag Dataset. Anyways each approach uses Short-time Fourier Transform (SFTF)[9] and it allows us to see how different frequencies change over time.

FIGURE 6
RAW WAVE INPUT



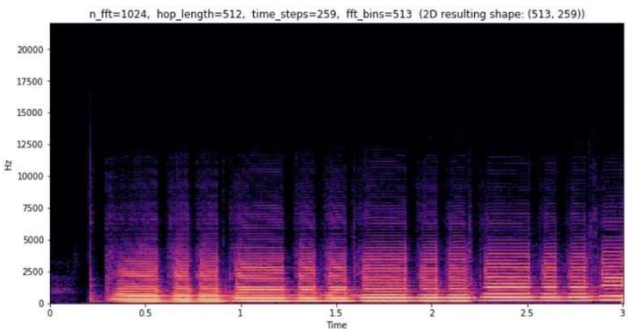


FIGURE 7 :MEL SPECTROGRAM INPUT

The feature vector that can be used consist of:

- ZCR: Zero-crossing rate
- Chroma: Chromagram 1
- Chroma 2: Chromagram 2
- MFCC: Mel-frequency cepstral coefficients
- Loudness: Loudness
- Energy: Energy

And both audio and video inferences can finally be cumulated through fuzzy inference.

- Priority based Scheduling: I can devise an algorithm such that I can efficiently assign computation task to particular video stream based on the probability that in future the violence can happen.

I devise a priority queue which has stores probability of violent activities. First, I initialize probability of each video stream to be 1/(Number of Video Streams) , and then upon any detection of violent activity, I update probability by,

$$P' = \min\left(1, P + ((isDetected) * e^{\frac{-confidence}{Number\ of\ Video\ Streams}}\right)$$

Here, isDetected is an integer variable which can be only -1 upon no detection or 1 upon detection. The confidence coefficient is updated after drawing inference from every CCTV is just the absolute sum of difference of previous confidence and output of the LSTM.

$$Initialize\ Confidence_i \leftarrow 0$$
$$Confidence' = |Confidence - \sum_i output_i|$$

And this is the scheduling queue according to which the video streams should be computed. Now the main topic of experimentation is to find the optimal time of updating the probabilities. If it's too small, the overhead of updating priority queue would add up to the delay. If the time of updating is too large, there is a possibility that a detected violent event is no taken under consideration. I can also make a robust pipeline for this. If possible, use multi-threading to achieve partial parallelism.

.

$$Frame\ Input \rightarrow Preprocessing \rightarrow Inference \rightarrow Update\ Confidence$$
$$Frame\ Input \rightarrow Preprocessing \rightarrow Inference \rightarrow Update\ Confidence$$
$$Frame\ Input \rightarrow Preprocessing \rightarrow Inference \ldots.$$

VI . CONCLUSION

In this work I implemented and experimented with various deepto predict violence in video data, I found our implementation to deal well with this task even though our GPU power was relatively low. I found the smart data preprocessing of the video's frames play an important factor as well as some of the training parameters such as: CNN network, learning rate and data augmentation. Looking forward to more complex violence scenarios and appliances it will takeresearchers to find creative solutions for data collection, advance generalization techniques and real-time optimizations. We also made a complementary app made in flutter for detecting violence over the CCTV or IP Webcam video streams using RTSP Protocol. Though the distributed computing cuts times in many folds still the system remains pseudo real time rather than real time while drawing inference. Though I am hopeful two-three papers down the line and we will achieve real time inference milestone, while maintaining the same precision. In conclusion, the development of a real-time violence detection system utilizing images uploading and CNN algorithms holds immense potential for enhancing public safety and security in various environments. Through the integration of advanced machine learning techniques and real-time video analysis, such systems offer a proactive approach to identifying and responding to potential threats promptly. Despite the advancements made in this field, there remain several areas for future research and development to further improve the effectiveness and reability of violence detection system.

REFERENCES

[1] Nandini Bagga,Gajan Singh, Balamurugan Balusamy, Ajay Shanker Singh(2022)"Violence Detection in Real Life Videos using Convolutional Neural Network".(Published in: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE))

[2] Silva Deena J, Md. Tabil Ahammed, Udaya Mouni Boppana, Maharin Afroj, Sudipto Ghosh, Sohaima Hossain(2022)"Real-time based Violence Detection from CCTV Camera using Machine Learning Method".(Published in: 2022 International Conference on Industry 4.0 Technology (I4Tech))

[3] Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2017) Action recognition in video
sequences using deep bi-directional LSTM with CNN features.(Published in: IEEE Access(Volume: 6))

[4] Yu Zhao and Rennong Yang and Guillaume Chevalier and Maoguo Gong (2017). "Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors".(Published in: Hindawi, volume 2018)

[5] Sudhakaran, Swathikiran, and Oswald Lanz. "Learning to detect violent videos using convolution long short-term memory." Advanced Video and Signal Based Surveillance(AVSS), 2017 14th IEEE International Conference on, pp. 1-6. IEEE, 2017.

[6] Tal Hassner, Yossi Itcher, and Orit Kliper-Gross. Violent flows: Realtime detection of violent crowd behavior. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society
Conference on, pages 1–6, IEEE, 2012