Enhancing Public Safety Through Intelligent Video Analytics For Violence Detection

Er.R.Sivaranjini[1], S.Sakthishree[2], M.Nishanthini[3], R.Anusuya[4], A.R.Rajashri[5]
[1]Assistant Professor,Dept of Computer Science and Engineering
[2, 3, 4, 5]Dept of Computer Science and Engineering
[1, 2, 3, 4, 5] Krishnasamy College of Engineering and Technology,
Cuddalore, Tamil Nadu, India.

Abstract- *Surveillance cameras are becoming more prevalent worldwide thanks to advancements in digital video technology, but the sheer amount of footage generated makes it tough for humans to analyze in real-time. Even manual efforts can lead to delays in spotting events. However, with the rise of machine learning, tasks like automatically detecting violence in videos have become more feasible.*

Our study delves into using intelligent networks that use 3D convolutions to understand the changing relationships between people and objects over time, capturing both the spatial and temporal aspects of the data. We also make use of a pre-trained model for action recognition to enhance the efficiency and accuracy of violence detection in surveillance videos. To validate our methods, we assess them on various public datasets containing diverse and challenging video content.

Our experimental results show that the proposed method achieves a substantial improvement in detection performance, increasing accuracy from 75% in existing systems to 91%. This demonstrates the effectiveness of our approach in handling complex real-world surveillance scenarios while maintaining a lightweight and efficient model architecture suitable for real-time applications.

Keywords- Violence detection,Video analyze, Action recognition

I. INTRODUCTION

In today's society, surveillance and security cameras are strategically placed in public areas to monitor events and human behavior, enhancing public safety and deterring crime. The footage captured by these cameras serves as crucial evidence in legal proceedings. Identifying and addressing anomalies like violence swiftly is imperative for military and law enforcement agencies to maintain security and reduce crime rates. However, the vast amount of video data generated by surveillance cameras on a daily basis, coupled with the infrequency of violent incidents compared to routine activities, makes manual monitoring impractical and error-prone. Thus, there's a pressing need for automated and effective methods to detect abnormal or violent behaviors, particularly in surveillance videos.

Human Activity Recognition (HAR) through video classification has garnered significant attention in recent years, mirroring the objectives of violence detection. These methods utilize sensor data to understand human actions, ranging from simple tasks like standing to more complex activities like cooking or conversing. Early HAR techniques focused on identifying and tracking human body parts using image descriptors like Histogram of Oriented Gradients (HOG) or Histogram of OrientedOptical Flow (HOF), often reliant on favorable lighting conditions and clear visibility. The advent of depth cameras has led to the development of algorithms leveraging depth measurements from devices such as Microsoft Kinect or Intel RealSense for HAR. These depth sensors offer the advantage of real-time skeletal tracking through Software Development Kits (SDKs), enabling the detection and description of human actions based on tracked joint coordinates over time. Despite advancements, modern depth sensors still exhibit considerable noise in their measurements, a challenge that researchers continue to address through innovative algorithms for HAR, often combining color and depth sensor data for improved accuracy.

II. LITERATURE REVIEW

A REAL TIME CRIME SCENE INTELLIGENT VIDEO SURVEILLANCE SYSTEMS IN VIOLENCE DETECTION FRAMEWORK USING DEEP LEARNING TECHNIQUES

Kishan Bhushan Sahay,Dr. Bhuvaneswari Balachander,Dr. B. Jagadeesh

Surveillance system research is now experiencing great expansion. Surveillance cameras put in public locations such as offices, hospitals, schools, roads, and other locations can be utilised to capture important activities and movements

for event prediction, online monitoring, goal-driven analysis, and intrusion detection. This research proposed novel technique in detecting crime scene video surveillance system in real time violence detection using deep learning architectures. Here the aim is to collect the real time crime scene video of surveillance system and extract the features using spatio temporal (ST) technique with Deep Reinforcement neural network (DRNN) based classification technique. The input video has been processed and converted as video frames and from the video frames the features has been extracted and classified. Its purpose is to detect signals of hostility and violence in real time, allowing abnormalities to be distinguished from typical patterns. To validate our system's performance, it is trained as well as tested in large-scale UCF Crime anomaly dataset. The experimental results reveal that the suggested technique performs well in real-time datasets, with accuracy of 98%, precision of 96%, recall of 80%t, and F-1 score of 78%.

EFFICIENT VIOLENCE DETECTION IN SURVEILLANCE

Romas Vijeikis,Vidas Raudonis andGintaras Dervinis

Intelligent video surveillance systems are rapidly being introduced to public places. The adoption of computer vision and machine learning techniques enables various applications for collected video features; one of the major is safety monitoring. The efficacy of violent event detection is measured by the efficiency and accuracy of violent event detection. In this paper, we present a novel architecture for violence detection from video surveillance cameras. Our proposed model is a spatial feature extracting a U-Net-like network that uses MobileNet V2 as an encoder followed by LSTM for temporal feature extraction and classification. The proposed model is computationally light and still achieves good results—experiments showed that an average accuracy is 0.82 ± 2% and average precision is 0.81 ± 3% using a complex real-world security camera footage dataset based on RWF-2000.

PROBLEM STATEMENT

In automated video surveillance systems, promptly and accurately identifying violence is essential for upholding public safety. The current approach involves several steps. Firstly, relevant features are extracted from video frames, such as color histograms, motion vectors, and spatial-temporal descriptors. Then, deep learning architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or Convolutional Recurrent Neural Networks (CRNNs) are applied to learn hierarchical features from raw pixel data and capture temporal relationships within video sequences. Annotated datasets containing labeled instances of violence and non-violence are crucial for training these models, and techniques like data augmentation are utilized to increase dataset diversity and improve model generalization. This comprehensive strategy aims to achieve both speed and accuracy in violence detection, enhancing the efficacy of automated video surveillance systems in safeguarding public security.

III. PROPOSED SYSTEM

Incorporating an email notification feature into automated video surveillance systems upon detecting violence is a vital enhancement. Once violence is identified using the established methods, a trigger mechanism activates the email notification process. This entails creating a module within the surveillance system to compose and dispatch emails to designated authorities or security personnel. These emails can contain pertinent information such as the incident's location, timestamps, and possibly snapshots or video excerpts for immediate evaluation. It's essential to implement secure and efficient protocols for sending emails to ensure the reliability and integrity of the notification system. This additional functionality reinforces the proactive nature of automated surveillance systems, facilitating prompt intervention and response to potential security risks or violent occurrences.

ADVANTAGES

- Violence detection systems offer the capability to recognize potentially violent situations in real-time or near-real-time, facilitating prompt intervention by security or law enforcement authorities. This proactive approach aids in preventing escalation and reducing harm.
- Integrating violence detection systems into public areas, transportation centers, educational institutions, and other vital infrastructure sites can bolster overall security and safety for individuals and communities, providing an added layer of defense against violence threats.
- These systems streamline the monitoring and analysis of video feeds or audio recordings for signs of violence, thereby relieving security personnel of this task and enabling them to focus on other crucial responsibilities.
- The existence of violence detection systems serves as a deterrent against violent behavior, as individuals are aware of being under surveillance by automated systems, potentially dissuading them from engaging in violent acts and fostering a safer environment.

PROPOSED ARCHITECTURE DESIGN

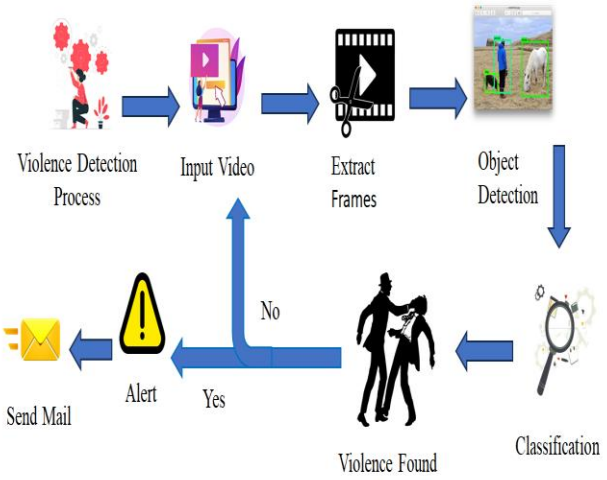


Figure 1: System Architecture

MOBILENETV2

MobileNetV2 is a convolutional neural network (CNN) model tailored for lightweight and efficient deep learning tasks, particularly suited for mobile and embedded systems. It follows the footsteps of the original MobileNet, developed by Google researchers.

The primary aim of MobileNetV2 is to attain top-tier accuracy on computer vision tasks while drastically reducing model size and computational complexity compared to bulkier models like VGG or ResNet. This makes it ideal for constrained environments such as mobile devices and IoT gadgets where resources like computation power and memory are limited.

Here are the key characteristics of MobileNetV2:

1. Inverted Residuals with Linear Bottlenecks: MobileNetV2 introduces a unique architecture termed inverted residuals, comprising a lightweight bottleneck layer followed by a linear activation function. This design cuts down on parameters while maintaining representational capacity.
2. Linear Bottlenecks: Linear bottlenecks aim to minimize non-linearity in the model, aiding in better optimization during training.
3. Expansion and Contraction: MobileNetV2 adjusts the number of channels in each layer dynamically using expansion and contraction factors. This helps find a balance between model complexity and computational efficiency.
4. Inverted Residuals and Linear Bottlenecks with Shortcut Connections: Similar to ResNet, MobileNetV2 incorporates residual connections to ease gradient flow and address the vanishing gradient issue during training.
5. Depthwise Separable Convolution: Like its predecessor, MobileNetV2 employs depthwise separable convolutions, splitting standard convolutions into depthwise and pointwise convolutions to reduce parameters and computations.
6. Efficient Architecture: MobileNetV2 is designed with efficiency in mind, optimizing both architecture and training processes to achieve high performance with minimal computational resources.

MobileNetV2 has found widespread adoption in various applications, including image classification, object detection, and semantic segmentation, owing to its compact size, speed, and competitive accuracy compared to larger CNN architectures.

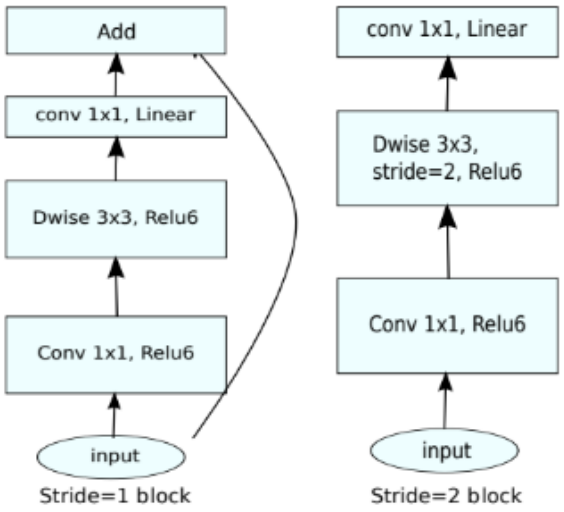


Figure 2: Control Flow

LSTM

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) architecture, crafted to address the shortcomings of traditional RNNs in grasping and maintaining long-term connections in sequential data. Unlike traditional RNNs, which may falter in learning from distantly spaced sequences, LSTMs are furnished with specialized memory cells, enabling them to retain information over extended durations.

Here's a simplified explanation of how LSTMs function:

1. **Cell State (Ct)**: Think of the cell state as a conveyor belt that carries information across different time steps. Unlike traditional RNNs where the hidden state is updated at each step, the cell state in LSTMs is modified selectively through gates, aiding in preserving information over lengthy sequences.
2. **Forget Gate (ft)**: The forget gate decides which details from the previous cell state should be discarded or remembered. By considering the previous hidden state and current input, it assigns values between 0 and 1 to each element in the cell state. A value close to 1 means "retain this information," while close to 0 means "discard it."
3. **Input Gate (it)**: This gate determines which new information to incorporate into the cell state. Comprising a sigmoid layer and a tanh layer, it selects values to be updated and generates a vector of potential new information to add to the cell state.
4. **Output Gate (ot)**: The output gate regulates the flow of information from the cell state to the output based on the current input and previous hidden state. It governs what information is passed to the next time step.
5. **Hidden State (ht)**: The hidden state acts as the "memory" of the LSTM cell, retaining information about the current time step. Calculated based on the current input, previous hidden state, and current cell state, it's utilized for predictions or generating outputs at each time step.

LSTMs excel in tasks involving sequential data, such as language processing, speech recognition, and time series forecasting, where capturing enduring connections is pivotal. Their capacity to selectively preserve and update information over time renders them invaluable in modeling and deciphering sequential patterns.

IV. MODULES

DATASET PREPARATION

The dataset preparation module stands as a fundamental aspect of any machine learning endeavor, laying the groundwork for model development and assessment. This module encompasses a series of steps, including data gathering, preprocessing, labeling, and validation, all aimed at ensuring the dataset's quality and coherence.

Upon collection, raw data often necessitates preprocessing to guarantee uniformity, reliability, and compatibility with the chosen machine learning techniques. This involves tasks like refining noisy data, managing missing values, standardizing formats, and scaling features.

Throughout the dataset preparation journey, validation and quality assurance play pivotal roles in pinpointing and rectifying errors, biases, or incongruities. Approaches such as cross-validation, outlier identification, and statistical scrutiny serve to uphold the trustworthiness and resilience of the dataset.

MODEL CREATION

The process encompasses several critical stages, beginning with problem definition and algorithm choice. Data scientists select the most suitable algorithm or model architecture based on the problem type (classification, regression, etc.) and data characteristics to attain the desired results.

Once the algorithm is chosen, the subsequent step involves crafting the model architecture. This entails determining the layer count, activation functions, and otherhyperparameters dictating the model's behavior. The architecture should be meticulously crafted to strike a balance between complexity and efficiency, ensuring the model adeptly captures underlying data patterns.

Following architecture definition, the model undergoes training using labeled data. Throughout this phase, the model learns to make predictions by fine-tuning its parameters to minimize a predefined loss function. This typically involves employing optimization techniques like stochastic gradient descent (SGD) or its variations.

Throughout the model creation journey, it's imperative to iterate and experiment with diverse architectures, hyperparameters, and training methodologies to enhance performance and generalization. Moreover, rigorous validation and testing procedures should be adhered to, guaranteeing the model's reliability across various datasets and scenarios.

VIDEO ANALYZE MODULE

The video analysis module plays a crucial role in computer vision systems by extracting valuable insights from video streams. This module encompasses a range of techniques and algorithms for processing, comprehending, and interpreting video data, leading to applications in diverse fields like surveillance, security, entertainment, healthcare, and beyond.

The process begins with acquiring video data from various sources such as video files, live camera feeds, or streaming platforms. It involves functionalities for capturing, decoding, and preprocessing video frames to prepare them for analysis.

Subsequently, video frames undergo sequential processing using computer vision algorithms to extract pertinent information. This may entail tasks such as identifying objects, tracking their movement, segmenting scenes, recognizing actions, faces, and comprehending the overall scene context.

Object detection algorithms are employed to pinpoint and localize objects of interest within each frame, while object tracking algorithms monitor the movement of these objects across consecutive frames. This facilitates tasks such as surveillance, traffic monitoring, and activity analysis, among others.

VIOLENCE DETECTION

The violence detection module is a specialized segment within video analysis systems, dedicated to identifying and marking instances of violent behavior in video streams. It holds significant importance in various fields like security surveillance, law enforcement, crowd management, and public safety.

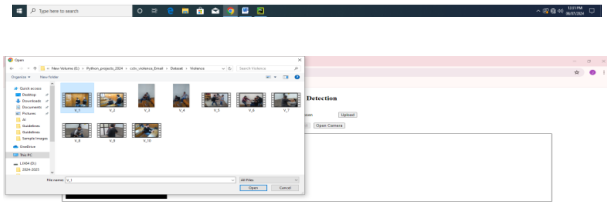
To begin its operation, the violence detection module initially focuses on recognizing and tracking relevant objects or individuals within video frames. This includes identifying human subjects and continuously monitoring their movements and interactions in real-time.
Moreover, the module emphasizes real-time processing capabilities to enable swift identification and response to violent incidents. This is achieved through the use of efficient algorithms, parallel processing methods, and hardware acceleration, ensuring minimal latency performance.

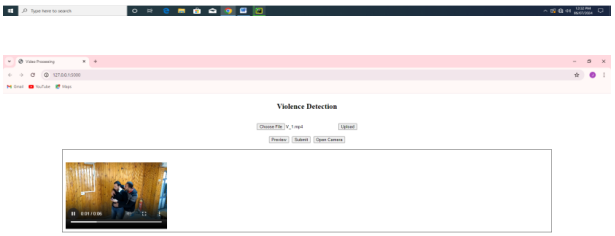
NOTIFICATION MODULE

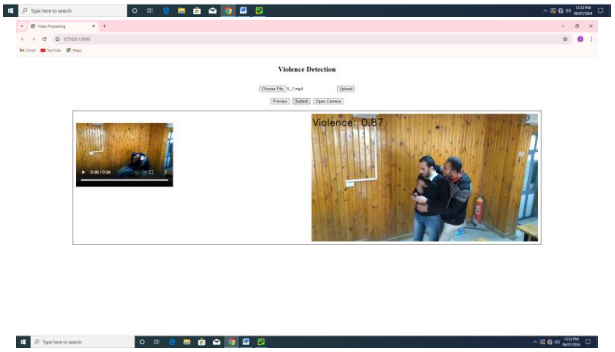
The notification module serves as an essential element in systems geared towards monitoring and analyzing data streams, delivering timely alerts and notifications to users or stakeholders in response to predefined criteria or events. This module holds significant importance across a range of applications, including security surveillance, real-time monitoring, healthcare, and industrial automation.

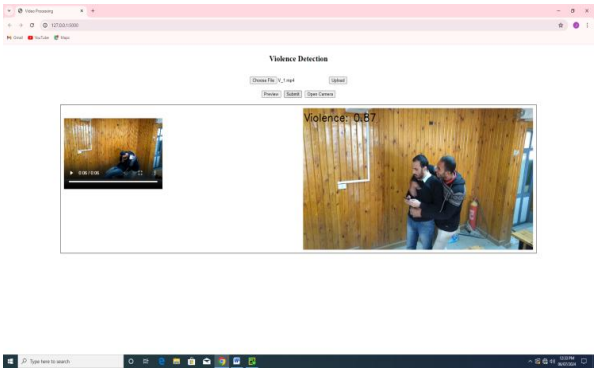
Operating within the system, the notification module continuously observes incoming data streams for particular events or conditions of significance. These may encompass anomalies, breaches of thresholds, critical incidents, or predetermined triggers established according to user-defined rules or algorithms.

V. RESULT









VI. CONCLUSION

This work demonstrates achieving high accuracy and robustness in real-time violence detection from videos with limited data and computational resources using transfer learning. The system employs the MobileNet architecture for violence detection. The proposed base model is evaluated on a standard dataset, surpassing previous research with an accuracy of 91%. Additionally, it exhibits faster performance compared to prior approaches. Despite these achievements, there's room for improvement by curating a comprehensive, balanced dataset from diverse video sources to enhance the sophistication of violence detection, focusing on identifying the violent actions themselves rather than solely detecting the presence or absence of violence.

VII. FUTURE ENHANCEMENTS

Improved Accuracy and Efficiency:

Future research aims to enhance the accuracy and efficiency of violence detection algorithms through advancements in deep learning architectures, feature extraction techniques, and data augmentation strategies. This includes developing models that can effectively differentiate between various types of violent behaviors and contextual factors.

Multimodal Fusion:

Integrating multiple modalities such as video, audio, text, and sensor data can improve the robustness and reliability of violence detection systems. Future research will explore techniques for multimodal fusion to leverage complementary information from different sources and enhance overall performance.

Real-time and Edge Computing:

There is a growing demand for real-time violence detection systems that can analyze video streams in real-time and provide immediate responses to potential threats. Future developments will focus on optimizing algorithms for edge computing devices, enabling low-latency processing and decentralized deployment in smart cameras, drones, and IoT devices.

Privacy-preserving Techniques:

As concerns about privacy and data security increase, future violence detection systems will incorporate privacy-preserving techniques to ensure compliance with regulations and protect individuals' privacy rights. This includes developing algorithms that can analyze video data while preserving the anonymity of individuals and sensitive information.

REFERENCES

[1] D. Ganesh, R. R. Teja, C. D. Reddy, and D. Swathi, “Human Action Recognition based on Depth maps, Skeleton and Sensor Images using Deep Learning,” in 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), IEEE, Oct. 2022,pp.1–8doi:10.1109/GCAT55367.2022.9971982.

[2] M. DALLEL, V. HAVARD, D. BAUDRY, and X. SAVATIER, “In HARD - Industrial Human Action Recognition Dataset in the Context of Industrial Collaborative Robotics,” in 2020 IEEE International Conference on Human-Machine Systems (ICHMS), IEEE, Sep. 2020, pp. 1–6. doi: 10.1109/ICHMS49158.2020.9209531.

[3] J.-S. Kim, “Efficient Human Action Recognition with Dual-Action Neural Networks for Virtual Sports Training,” in 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), IEEE, Oct. 2022, pp. 1–3. doi: 10.1109/ICCE-Asia57006.2022.9954758.

[4] P. Le Noury, R. Polman, M. Maloney, and A. Gorman, “A Narrative Review of the Current State of Extended Reality Technology and How it can be Utilised in Sport,” Sports Medicine, vol. 52, no. 7, pp. 1473–1489, Jul. 2022, doi: 10.1007/s40279-022-01669-0.

[5] T. Hassner, Y. Itcher, and O. Kliper-Gross. Violent flows: Real-time detection of violent crowd behavior. In CVPR Workshops, June 2012.

[6] Serrano Gracia I, Deniz Suarez O, Bueno Garcia G, Kim TK. Fast fight detection. PLoS One. 2015 Apr10;10(4):e0120448.doi:10.1371/journal.pone.0120448. PMID: 25860667; PMCID: PMC4393294.

[7] Deniz, Oscar & Serrano Gracia, Ismael &Bueno, Gloria & Kim, Tae-Tyun. (2014). Fast violence detection in video. VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications. 2.

[8] Hassner, T., Itcher, Y., &Kliper-Gross, O. (2012). Violent flows: Real-time detection of violent crowd behavior. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 1-6.

[9] Sudhakaran, Swathikiran&Lanz, Oswald. (2017). Learning to detect violent videos using convolutional long short-term memory. 1-6. 10.1109/AVSS.2017.8078468.

[10] Li, Ji& Jiang, Xinghao& Sun, Tanfeng&xu, ke. (2019). Efficient Violence Detection Using 3D ConvolutionalNeural Networks.1-8. 10.1109/AVSS.2019.8909883.

[11] Howard, Andrew & Zhu, Menglong& Chen, Bo &Kalenichenko, Dmitry & Wang, Weijun&Weyand, Tobias &Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

[12] Benjamin Lindemann, Timo Müller, HannesVietz, Nasser Jazdi, Michael Weyrich,A survey on long short-term memory networks for time series prediction, ProcediaCIRP,Volume 99, 2021, Pages650-655, ISSN 22128271,https://doi.org/10.1016/j.procir.2021.03.088.