Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN: 2581 - 611X)

VOLUME:8, ISSUE:1, FEBRUARY 2025, PAPER NUMBER:30

Real Time Violence Detection Using MobileNetV2

Dr.R Deeptha
dept. of. Information Technology
deepthar@srmist.edu.in

Varun.P
dept. of Information Technology
vp3530@srmist.edu.in

Padmashree Shreya V
dept. of. Information Technology
pv3430@srmist.edu.in

Santhosh Kumar C
dept.of Information Technology
sc9739@srmist.edu.in

***Abstract*—** The pervasive occurrence of violent activities worldwide poses a significant threat to personal safety and societal stability. Various strategies, including the deployment of surveillance systems, have been employed to mitigate these activities. The development of surveillance systems capable of autonomously detecting violent incidents and issuing timely warnings or alerts holds immense importance.

This proposed system entails a structured sequence of procedures. Initially, it identifies human presence within video frames, followed by the extraction of frames likely to contain violent behavior while discarding irrelevant ones. Subsequently, a trained model discerns violent conduct, saving relevant frames as images. If feasible, these images undergo enhancement to facilitate face detection. Accompanied by essential details such as time and location, these enhanced images are then dispatched as alerts to relevant authorities.

Central to this approach is a deep learning framework utilizing Convolutional Neural Networks (CNN) to identify violence within videos. However, CNN alone may exhibit drawbacks, including prolonged computation times and reduced accuracy. To address these limitations, a pre-trained model, MobileNet, is leveraged for its superior accuracy and as a foundation for the overall system.

Alerts are efficiently disseminated to concerned authorities through the Telegram application.

***Keywords*—** The pervasive occurrence of violent activities worldwide poses a significant threat to personal safety and societal stability. Various strategies, including the deployment of surveillance systems, have been employed to mitigate these activities. The development of surveillance systems capable of autonomously detecting violent incidents and issuing timely warnings or alerts holds immense importance.

The pervasive occurrence of violent activities worldwide poses a significant threat to personal safety and societal stability. Various strategies, including the deployment of surveillance systems, have been employed to mitigate these activities. The development of surveillance systems capable of autonomously detecting violent incidents and issuing timely warnings or alerts holds immense importance.

This proposed system entails a structured sequence of procedures. Initially, it identifies human presence within video frames, followed by the extraction of frames likely to contain violent behavior while discarding irrelevant ones. Subsequently, a trained model discerns violent conduct, saving relevant frames as images. If feasible, these images undergo enhancement to facilitate face detection. Accompanied by essential details such as time and location, these enhanced images are then dispatched as alerts to relevant authorities.

Central to this approach is a deep learning framework utilizing Convolutional Neural Networks (CNN) to identify violence within videos. However, CNN alone may exhibit drawbacks, including prolonged computation times and reduced accuracy. To address these limitations, a pre-trained model, MobileNet, is leveraged for its superior accuracy and as a foundation for the overall system.

Alerts are efficiently disseminated to concerned authorities through the Telegram application. The pervasive occurrence of violent activities worldwide poses a significant threat to personal safety and societal stability. Various strategies, including the deployment of surveillance systems, have been employed to mitigate these activities. The development of surveillance systems capable of autonomously detecting violent incidents and issuing timely warnings or alerts holds immense importance.

This proposed system entails a structured sequence of procedures. Initially, it identifies human presence within video frames, followed by the extraction of frames likely to contain violent behavior while discarding irrelevant ones. Subsequently, a trained model discerns violent conduct, saving relevant frames as images. If feasible, these images undergo enhancement to facilitate face detection. Accompanied by essential details such as time and location, these enhanced images are then dispatched as alerts to relevant authorities.

Central to this approach is a deep learning framework utilizing Convolutional Neural Networks (CNN) to identify violence within videos. However, CNN alone may exhibit drawbacks, including prolonged computation times and reduced accuracy. To address these limitations, a pre-trained model, MobileNet, is leveraged for its superior accuracy and as a foundation for the overall system.

Alerts are efficiently disseminated to concerned authorities through the Telegram application.

II. LITERATURE REVIEW

2.1 OVERVIEW

Recently proposed methods for violence detection can be classified into three categories: visual-based approach, audio-based approach, and hybrid approach.

Visual-Based Approach: This approach retrieves visual information and represents it as relevant features. It includes local features (such as position, velocity, form, and color) and global features (like average speed, region occupancy, and interactions between objects and backdrop).

Audio-Based Approach: Violence classification relies on audio data in this approach. It utilizes hierarchical techniques based on Gaussian mixture models and Hidden Markov models to distinguish sounds like gunshots, explosions, and automobile braking.

Hybrid Approach: This method merges visual and audio characteristics. Techniques under this approach utilize flame and blood detection, record the degree of motion, and identify typical sounds of violent occurrences. For instance, the CASSANDRA system combines motion features from video with scream-like cues from audio.

1.“J. K. Aggarwal and M. S. Ryoo, “Human activity analysis: A review,” ACM Comput. Surv., vol. 43, no. 3, 2011, doi: 10.1145/1922649.1922653. ”Human activity recognition is an important area of computer vision research. Its applications include surveillance systems, patient monitoring systems, and a variety of systems that involve interactions between persons and electronic devices such as human-computer interfaces. Most of these applications require an automated recognition of high-level activities, composed of multiple simple (or atomic) actions of persons. This article provides a detailed overview of various state-of-the-art research papers on human activity recognition. We discuss both the methodologies developed for simple human actions and those for high-level activities. An approach-based taxonomy is chosen that compares the advantages and limitations of each approach.

Recognition methodologies for an analysis of the simple actions of a single person are first presented in the article.Space-time volume approaches and sequential approaches that represent and recognize activities directly from input images are discussed. Next, hierarchical recognition methodologies for high-level activities are presented and compared. Statistical approaches, syntactic approaches, and description-based approaches for hierarchical recognition are discussed in the article. In addition, we further discuss the papers on the recognition of human-object interactions and group activities. Public datasets designed for the evaluation of the recognition methodologies are illustrated in our article as well, comparing the methodologies' performances. This review will provide the impetus for future research in more productive areas.

2.“G. Cheng, Y. Wan, A. N. Saudagar, K. Namuduri,and B. P. Buckles, “Advances in Human Action Recognition: A Survey,” no. February, 2015, [Online]. Available: http://arxiv.org/abs/1501.05964.” Human action recognition is an important application domain in computer vision. Its primary aim is to accurately describe human actions and their interactions from a previously unseen data sequence acquired by sensors. The ability to recognize, understand and predict complex human actions enables the construction of many important applications such as intelligent surveillance systems, human-computer interfaces, health care, security and military applications. In recent years, deep learning has been given particular attention by the computer vision community. This paper presents an overview of the current state-of-the-art in action recognition using video analysis with deep learning techniques. We present the most important deep learning models for recognizing human actions, analyze them to provide the current progress of deep learning algorithms applied to solve human action recognition problems in realistic videos highlighting their advantages and disadvantages.

Based on the quantitative analysis using recognition accuracies reported in the literature, our study identifies state-of-the-art deep architectures in action recognition and then provides current trends and open problems for future works in this filed.

3.“S. A. R. Abu-Bakar, “Advances in human action recognition: An updated survey,IET Image Process., vol. 13, no. 13, pp. 2381–2394, 2019, doi: 10.1049/ietipr.2019.0350”. Research in human activity recognition (HAR) has seen tremendous growth and continuously receiving attention from both the Computer Vision and the Image Processing communities. Due to the existence of numerous publications in this field, undoubtedly, there have been a number of review papers on this subject that categorise these techniques. Many of the recent works have started to tackle more challenging problems and these proposed techniques are addressing more realistic real-world scenarios. Conspicuously, an updated survey that covers these methods is timely due. To simplify the categorisation, this study takes a two-layer hierarchical approach. At the top level, the categorisation is based on the basic process flow of HAR, i.e. input data-type, features-type, descriptor-type, and classifier-type. At the second layer, each of these components is further subcategorised based on the diversity of the proposed methods. Finally, a remark on the coming popularity of deep learning approach in this field is also given.

4. "P.Guo, Z. Miao, Y. Shen, W. Xu, and D. Zhang, "Continuous human action recognition in real time," Multimed. Tools Appl., vol. 68, no. 3, pp. 827–844, 2014, doi: 10.1007/s11042-012-1084-2. "This paper discusses the task of continuous human action recognition. By continuous, it refers to videos that contain multiple actions which are connected together. This task is important to applications like video surveillance and content based video retrieval. It aims to identify the action category and detect the start and end key frame of each action. It is a challenging task due to the frequent changes of human actions and the ambiguity of action boundaries. In this paper, a novel and efficient continuous action recognition framework is proposed. Our approach is based on the bag of words representation. A visual local pattern is regarded as a word and the action is modeled by the distribution of words. A generative translation and scale invariant probabilistic Latent Semantic Analysis model is presented. The continuous action recognition result is obtained frame by frame and updated from time to time. Experimental results show that this approach is effective and efficient to recognize both isolated actions and continuous actions.

5."A. Jalal, M. Uddin, and T. S. Kim, "Depth videobased human activity recognition system using translation and scaling invariant features for life logging at smart home," IEEE Trans. Consum. Electron., vol. 58, no. 3, pp. 863– 871, 2012, doi: 10.1109/TCE.2012.6311329. [10] J. Hu and N. V. Boulgouris, "Fast human activity recognition based on structure and motion," Pattern Recognit. Lett., vol. 32, no. 14, pp. 1814–1821, 2011, doi: 10.1016/j.patrec.2011.07.013." Video-based human activity recognition systems have potential contributions to various applications such as smart homes and healthcare services. In this work, we present a novel depth video-based translation and scaling invariant human activity recognition (HAR) system utilizing R transformation of depth silhouettes. To perform HAR in indoor settings, an invariant HAR method is critical to freely perform activities anywhere in a camera view without translation and scaling problems of human body silhouettes. We obtain such invariant features via R transformation on depth silhouettes. Furthermore, in R transforming depth silhouettes, shape information of human body reflected in depth values is encoded into the features. In R transformation, 2D feature maps are computed first through Radon transform of each depth silhouette followed by computing 1D feature profile through R transform to get the translation and scaling invariant features. Then, we apply Principle Component Analysis (PCA) for dimension reduction and Linear Discriminant Analysis (LDA) to make the features more prominent, compact and robust. Finally, Hidden Markov Models (HMMs) are used to train and recognize different human activities. Our proposed system shows superior recognition rate over the conventional approaches, reaching up to the mean recognition rate of 93.16% for six typical human activities whereas the conventional PC and IC-based depth silhouettes achieved only 74.83% and 86.33% ,while binary silhouettes-based R transformation approach achieved 67.08% respectively. Our experimental results show that the proposed method is robust, reliable, and efficient in recognizing the daily human activities.

2.3 INFERENCE

Collectively, these literature reviews showcase the evolution of methodologies in human activity recognition and suspicious activities detection. The shift towards deep learning models is evident, emphasizing their effectiveness in handling complex tasks and addressing real-world challenges. The integration of deep learning techniques, hierarchical categorization approaches, and innovative frameworks reflects the ongoing efforts to enhance the accuracy, efficiency, and applicability of computer vision systems in diverse scenarios. state the units for each quantity that you use in an equation.

3.SYSTEM ANALYSIS

3.1 Existing System in Surveillance Activity Detection:

The current Violence surveillance system relies on CCTV cameras to identify criminals after a crime has been committed, rather than detecting suspicious activities beforehand. This reactive approach limits the effectiveness of law enforcement in preventing criminal acts and poses significant challenges in terms of data management and resource allocation. The existing system stores video data in the form of records, requiring continuous monitoring, which is a labor-intensive task.

3.2 Disadvantages of Existing Systems:

1. Reactive Approach: The surveillance system primarily serves as a tool for post-incident investigation rather than proactive crime prevention. This limits its utility in deterring criminal activities before they occur.

2. Data Storage and Management: Storing video datain the form of records necessitates significant storage space and can lead to data redundancy. Managing and analyzing this vast amount of data is labor-intensive and time-consuming.

3. Continuous Monitoring Requirement: The current system relies on continuous human monitoring of surveillance footage, which is resource-intensive and prone to errors due to human fatigue.

4. Delayed Response Time: Since suspicious activities are not detected in real-time, law enforcement agencies may respond to incidents with a delay, allowing

criminals to escape or carry out their activities unhindered.

5. Limited Efficiency in Crime Prevention: Without the ability to detect and intervene in suspicious activities before they escalate into criminal acts, the surveillance system's effectiveness in crime prevention is compromised.

3.3Proposed System:

The objective of this project was to develop an automated system capable of detecting violent incidents in surveillance footage autonomously. Leveraging machine learning, video segmentation, and anomaly detection techniques, the system analyzes video streams to identify frames containing potentially violent behavior, such as physical altercations or brandishing weapons. Implementation involved utilizing OpenCV for video processing, MobilenetV2 for feature extraction, and MTCNN for face detection. Additionally, a Convolutional Neural Network (CNN) was trained on annotated datasets to accurately classify violent activities. The system was implemented in Python, incorporating libraries such as OpenCV, MobilenetV2, MTCNN, and integrated with a Telegram bot for prompt alert generation to relevant authorities.

3.4 ADVANTAGES OF PROPOSED SYSTEM:

1. Real-Time Response: By detecting abnormaactivities in real-time, the proposed system enables immediate intervention by relevant stakeholders, potentially preventing crimes or mitigating their impact.

2. Proactive Security Measures: Alert generation allows for proactive security measures to be taken, such as dispatching law enforcement or security personnel to investigate suspicious behavior before it escalates into criminal activity.

3. Enhanced Situational Awareness: Alerts provide stakeholders with timely information about securitythreats or incidents, enabling them to make informed decisions and respond effectively to emerging situations.

4. Resource Optimization: By automating the alert generation process, the proposed system reduces the reliance on continuous human monitoring, freeing up resources that can be allocated to other critical tasks within the surveillance operation.

5. Scalability: The system can be scaled to monitor large areas or multiple locations simultaneously without significantly increasing the burden on human operators, ensuring comprehensive coverage and security.

6. Customizable Alerting Mechanisms: The system allows for alerts to be customized based on the specific requirements of different applications or stakeholders, ensuring that relevant information is delivered in a timely manner through notifications, alarms, or other preferred channels.

7. Integration with Existing Infrastructure: The proposed system can integrate seamlessly with existing surveillance infrastructure, leveraging investments in hardware and software while enhancing overall functionality and effectiveness.

8. Reduced Response Time: With real-time alertgeneration, stakeholders can respond to security threats more rapidly, minimizing the potential for damage or loss associated with criminal activities.

9. Improved Crime Prevention: By promptly notifying stakeholders of abnormal activities, the proposed system contributes to the deterrence and prevention of criminal acts, creating a safer environment for individuals and communities.

10. Data-driven Decision Making: Alerts generated by the system are based on data-driven analysis of surveillance footage, providing stakeholders with actionable insights that support informed decision-making and proactive security strategies.

4.1Overview:

The proposed system aims to enhance surveillance capabilities by detecting abnormal activities in real-time and generating alerts to notify relevant stakeholders. By leveraging advanced computer vision and machine learning algorithms, the system can analyze video streams from CCTV cameras and identify suspicious behavior, such as unauthorized access, loitering, or unusual movements. Alerts are then generated and sent to stakeholders via notifications, alarms, or other preferred channels, enabling proactive intervention to prevent or mitigate security threats.

SYSTEM ARCHITECTURE

5.1 PROPOSED SYSTEM:

The project aimed to develop an automated system capable of autonomously detecting violent incidents in surveillance footage, thereby enhancing security measures and enabling timely intervention by relevant authorities. Leveraging advancements in machine learning, video segmentation, and anomaly detection techniques, the system was designed to scrutinize video streams and pinpoint frames exhibiting potentially violent behavior, including physical altercations or the presence of weapons.

The implementation of the system was multifaceted, integrating various technologies and libraries to achieve its objectives effectively. Here's a detailed breakdown of the key components and methodologies employed:

The process involved several key steps:

1.Video Processing with OpenCV:

OpenCV, a popular computer vision library, was utilized for video processing tasks. This involved parsing surveillance footage into individual frames for analysis, enabling efficient manipulation and extraction of relevant information.

2.Feature Extraction with MobilenetV2:

MobilenetV2, a lightweight deep neural network architecture optimized for mobile and embedded applications, was employed for feature extraction from video frames. Its efficiency and accuracy make it well-suited for identifying meaningful patterns and features indicative of violent behavior within the footage.

3.Face Detection with MTCNN:

Multi-Task Cascaded Convolutional Neural Network (MTCNN) was utilized for face detection within the video frames. This enabled the system to focus on relevant regions of interest, such as individuals engaged in potentially violent activities, enhancing the accuracy and specificity of the detection process.

4 Training CNN for Classification:

A Convolutional Neural Network (CNN) was trained on annotated datasets containing examples of violent activities. By learning from labeled data, the CNN could accurately classify frames as either normal or exhibiting violent behavior, thus enabling precise detection and classification of incidents.

5 Integration with Telegram Bot for Alert Generation:

To ensure prompt and efficient communication with relevant authorities, the system was integrated with a Telegram bot. Upon detecting a potentially violent incident, the system would generate alerts containing pertinent details, such as the timestamp, location, and nature of the observed activity, and dispatch them to designated recipients via the Telegram messaging platform.

Moreover, upon detection of abnormal activities, thesystem promptly generated alerts in real-time. These alerts could be disseminated to relevant stakeholders through various channels, including notifications, alarms, or customized alerts based on specific application

SYSTEM MODULES

Data Collection Module:

This module is crucial as it forms the foundation of the system. It involves gathering data from various sources such as surveillance cameras, sensors, logs, or any other relevant sources. The collected data could be in different formats including images, videos, text logs, or time-series data from sensors. The process of data collection should ensure that it captures a diverse range of scenarios and activities to train the model effectively.

Data Preprocessing Module:

Raw data collected from different sources often requires preprocessing before feeding it into the model. This module involves tasks such as:

- Data Cleaning: Removing noise, outliers, or irrelevant information from the data.
- Data Transformation: Converting data into a suitable format for training, which might involve resizing images, normalizing pixel values, or encoding categorical variables.
- Feature Scaling: Scaling features to ensure that they have similar ranges, which can improve the convergence of optimization algorithms during training.
- Handling Missing Data: Dealing with missing values in the dataset through imputation or deletion strategies.
- Data Augmentation: Generating additional training examples by applying transformations such as rotation, flipping, or cropping to enhance the model's ability to generalize.

Feature Extraction Module:

This module is responsible for extracting meaningful features from the preprocessed data that can capture the relevant information for detecting suspicious activities. The choice of feature extraction techniques depends on the type of data being used:

For image data, convolutional neural networks (CNNs) are commonly used to automatically extract hierarchical features from images.

For sequential data such as time-series or sensor readings, techniques like Fourier transforms, wavelet transforms, or recurrent neural networks (RNNs) may be used to capture temporal patterns.

The goal is to transform the raw data into a feature representation that is rich in information and suitable for training the model

MobileNet V2:

The MobileNet architecture is primarily based on depth wise sep arable convolution, in which factors a traditional convolution into a depth wise convo lution followed by a pointwise convolution. The module presents a residual cell (has a residual/identity connection) with stride of 1, and a resizing cell with a stride of 2. From Figure 3.3, "conv" is a normal con volution, "dwese" is a depth wise separable convolution, "Relu6" is a ReLu activation function with a magnitude limitation, and "Linear" is the use of the linear function. Themainstrategies introduced in MobilenetV2 were linear bottleneck and inverted residual blocks. In the linear bottleneck layer, the channel dimension of input is ex panded to reduce the risk of information loss by nonlinear functions such as ReLU. It stems from the fact that information lost in some channels might be preserved in other 9 channels. The inverted residual block has a ("narrow"-"wide"-""narrow") structure in the channel dimension whereas a conventional residual block has a ("wide"- "narrow" "wide") one. Since skip connections are between narrow layers instead of wider ones, the memory footprint can be reduced

Training Module:

This module involves training the MOBILENETV2 model using labeled data. The training process includes several steps:

- Data Splitting: Splitting the labeled data into training, validation, and test sets to evaluate the model's performance.
- Loss Function: Defining a suitable loss function that measures the difference between the model's predictions and the ground truth labels.
- Optimization Algorithm: Selecting an optimization algorithm (e.g., Adam, RMSProp) to minimize the loss function and update the model parameters during training.
- Hyperparameter Tuning: Tuning hyperparameters such as learning rate, batch size, or dropout rate to improve the model's performance.
- Training Loop: Iteratively feeding batches of data into the model, computing the loss, and updating the model parameters using backpropagation until convergence or a predefined stopping criterion is reached.

ImageEnhancement:

ImageEnhancement is performed on the frames that are obtained as output. This is performed using the inbuilt functions provided by the Python Imaging Library(PIL). PIL offers extensive file format support, efficient presentation, and fairly powerful image processing capabilities. The Core Image Library is designed to provide quick access to data stored in several major pixel formats. It provides a solid foundation for common image processing tools. The brightness and colour of the ob tained output frames is increased by a factor of 2

Evaluation Module:

Once the model is trained, it needs to be evaluated on unseen data to assess its performance in detecting suspicious activities. This module involves:

- Performance Metrics: Calculating evaluation metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve (AUC) to measure the model's performance.
- Confusion Matrix: Constructing a confusion matrix to visualize the model's predictions compared to the ground truth labels and identify any patterns of misclassification.
- Cross-Validation: Performing cross-validation to assess the model's generalization ability and robustness to variations in the dataset.
- Error Analysis: Analyzing specific cases where the model fails to detect suspicious activities to identify potential weaknesses or areas for improvement.

Deployment Module:

Once the model has been trained and evaluated, it needs to be deployed in a production environment where it can continuously monitor and detect suspicious activities in real-time. This module involves:

Integration: Integrating the model into the existing infrastructure, which may include deploying it on servers, edge devices, or cloud platforms.

Scalability: Ensuring that the deployed model can handle varying levels of workload and scale to accommodate increasing data volumes or user demands.

Reliability: Implementing mechanisms for monitoring the model's performance, detecting failures or anomalies, and automatically recovering from errors to ensure uninterrupted operation.

Security: Implementing security measures to protect sensitive data and prevent unauthorized access to the deployed system.

Alerting Module:

When the model detects suspicious activity, it needs to trigger alerts to notify appropriate stakeholders so that timely action can be taken. This module involves:

- Alert Generation: Generating alerts in response to detected suspicious activities, which may include sending notifications via email, SMS, or triggering alarms in the surveillance system.
- Alert Prioritization: Prioritizing alerts based on the severity or urgency of the detected activity to ensure that critical incidents receive prompt attention.
- Alert Handling: Implementing mechanisms for handling alerts, such as routing them to designated personnel or systems for further investigation or response.

CONCLUSION

Our study has shed light on the significant role of advanced technologies in effectively addressing contemporary security challenges. Through the implementation of MobileNetV2 and MTCNN models, we have demonstrated the potential to empower security professionals and law enforcement agencies with powerful tools for detecting and responding to suspicious activities.

The utilization of deep learning, particularly through MobileNetV2 and MTCNN models, represents a paradigm shift in the field of surveillance and security. By leveraging the inherent capabilities of neural networks to learn complex patterns from data, we can significantly enhance the accuracy and efficiency of suspicious activity recognition systems. The findings of our study underscore the importance of embracing these advanced technologies to stay ahead of evolving threats and safeguard public safety.

Looking ahead, further research and development efforts in this area hold immense promise for enhancing surveillance capabilities and mitigating security risks across various settings. Continued advancements in deep learning algorithms, fueled by the availability of large-scale annotated datasets and computational resources, will further propel the field of suspicious activity recognition towards greater accuracy and efficiency.

In particular, the integration of multi-modal data sources, such as audio recordings and sensor data, presents an exciting avenue for future research. By incorporating additional contextual information, we can enhance the robustness and reliability of suspicious activity detection systems, enabling more accurate threat assessment and response.

Moreover, the ongoing evolution of deep learning techniques, including transfer learning, domain adaptation, and online learning, will continue to drive improvements in model performance and adaptability. By leveraging pre-existing knowledge and adapting to changing environments, MobileNetV2 and MTCNN models can remain effective and relevant in dynamic surveillance scenarios.

Ultimately, our study contributes to the broader efforts aimed at leveraging cutting-edge technologies to enhance public safety and security. By developing and deploying innovative solutions like MobileNetV2 and MTCNN models, we can create safer environments for individuals and communities, mitigating potential threats and ensuring the well-being of society as a whole.

Moving forward, it is imperative to foster collaboration between academia, industry, and government agencies to facilitate the translation of research findings into practical applications. By working together, we can harness the full potential of advanced technologies to address emerging security challenges and foster a safer and more secure future for all.

FUTURE ENHANCEMENTS

Our study has illuminated the significant role of advanced technologies in effectively addressing contemporary security challenges, particularly in the realm of violence detection. Through the implementation of MobileNetV2 and MTCNN models, we have showcased the potential to empower security professionals and law enforcement agencies with robust tools for detecting and responding to suspicious activities indicative of violence.

The utilization of deep learning, particularly through MobileNetV2 and MTCNN models, signifies a paradigm shift in the field of surveillance and security. By harnessing the inherent capabilities of neural networks to learn intricate patterns from data, we can markedly enhance the accuracy and efficiency of violence detection systems. Our findings underscore the critical importance of embracing these advanced technologies to proactively mitigate potential threats and safeguard public safety.

Looking ahead, further research and development endeavors in this area hold immense promise for augmenting surveillance capabilities and mitigating security risks across diverse settings. Continued advancements in deep learning algorithms, supported by the availability of large-scale annotated datasets and computational resources, will drive the field of violence detection towards greater precision and effectiveness.

In particular, the integration of multi-modal data sources, such as audio recordings and sensor data, presents an intriguing avenue for future exploration. By incorporating additional contextual information, we can bolster the resilience and accuracy of violence detection systems, enabling more nuanced threat assessment and response strategies.

Furthermore, the ongoing evolution of deep learning techniques, encompassing transfer learning, domain adaptation, and online learning, will continue to catalyze improvements in model performance and adaptability. By leveraging existing knowledge and adapting to evolving circumstances, MobileNetV2 and MTCNN models can remain versatile and pertinent in dynamic surveillance scenarios.

REFERENCES

[1] J. K. Aggarwal and M. S. Ryoo, “Human activity analysis: A review,” ACM Comput. Surv., vol. 43, no. 3, 2011, doi: 10.1145/1922649.1922653.

[2] G. Cheng, Y. Wan, A. N. Saudagar, K. Namuduri, and B. P. Buckles, “Advances in Human Action Recognition: A Survey,” no. February, 2015, [Online]. Available: http://arxiv.org/abs/1501.05964.

[3] S. A. R. Abu-Bakar, “Advances in human action recognition: An updated survey,” IET Image Process., vol. 13, no. 13, pp. 2381–2394, 2019, doi: 10.1049/ietipr.2019.0350.

[4] P. Guo, Z. Miao, Y. Shen, W. Xu, and D. Zhang, “Continuous human action recognition in real time,” Multimed. Tools Appl., vol. 68, no. 3, pp. 827–844, 2014, doi: 10.1007/s11042-012-1084-2.

[5] A. Jalal, M. Uddin, and T. S. Kim, “Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home,” IEEE Trans. Consum. Electron., vol. 58, no. 3, pp. 863– 871, 2012, doi:

10.1109/TCE.2012.6311329. [10] J. Hu and N. V. Boulgouris, “Fast human activity recognition based on structure and motion,” Pattern Recognit. Lett., vol. 32, no. 14, pp. 1814–1821, 2011, doi: 10.1016/j.patrec.2011.07.013.

[6] J. K. Aggarwal and M. S. Ryoo, “Human activity analysis: A review,” ACM Computing Surveys, vol. 43, no. 3, 2011. [Online]. Available: https://doi.org/10.1145/1922649.1922653.