Real Time Violence Detection Using Deep Learning

PROJECT

SUBMITTED TO

Sanil P Nair

In partial fulfillment of the requirements for the

THIRD SEMESTER MINI PROJECT

in

DATA ANALYTICS

Submitted by

SreeVishak	233025
Sidharth Mohan	233022
Shahad PK	233021

SCHOOL OF DIGITAL SCIENCE, DIGITAL UNIVERSITY OF KERALA

Technopark Phase IV, Thiruvananthapuram, KERALA-695317

Jan 2025

BONAFIDE CERTIFICATE

This is to certify that the report, *Real Time Violence Detection Using Deep Learning*, submitted by:

Name:	**Reg No:**
SreeVishak	233025
Siddharth Mohan	233022
Shahad P K	233021

This is a Bonafide record of the work carried out at Kerala University of Digital Sciences, Innovation and Technology under my supervision, which partially fulfills the requirements for the award of a Master of Science in Computer Science with specialization in Data Analytics.

Supervisor
Sanil P Nair
School of Digital Sciences
Digital University Kerala

DECLARATION

We, **SreeVishak, Siddharth Mohan, and Shahad P K**, students of Master of Science Computer Science with specialization in Data Analytics, hereby declare that the project report is substantially the result of our work. We affirm that the content of this report is the result of our efforts and research and that any external sources of information used have been duly acknowledged. We accept full responsibility for any instances of plagiarism found in this project.

Place: Trivandrum
Date: 27-07-2024

ACKNOWLEDGMENT

We would like to take this opportunity to express our sincere gratitude and appreciation to all those who have contributed to the successful completion of this project. This project would not have been possible without their backing, guidance, and encouragement. The first and foremost acknowledgment is to Sanil P Nair, our project supervisor for his invaluable support and profound guidance throughout the project period. His insights were integral in shaping the direction and methodology of this project. Moreover, we thank all Digital University Kerala staff and the administrative team for allowing us access to their resources and other necessary facilities that assisted in conducting the required research. We would also like to thank our friends and family members who have been a source of motivation and support during its development phase.

ABSTRACT

This project focuses on the development of a real-time violence detection system using computer vision and deep learning techniques for video surveillance applications. The main purpose is to design a deep learning model that is capable of classifying video frames into ”Violence” and ”Non-Violence”. This will help to maintain the safety and security of the public. The process begins with data collection, collected video dataset from Kaggle. Then the video dataset is converted into image frames using OpenCV. The dataset is split into training and test sets, followed by preprocessed into normalized and resized images. By using Convolutional Neural Network (CNN) architecture train the model to extract relevant features from images, and dropout layers are added to prevent overfitting. Used Adam optimizer as optimization algorithm and binary cross-entropy loss, and the performance was evaluated through accuracy and loss graphs. Also, a real-time application is developed by using Streamlit and OpenCV, where on-screen labels and confidence scores are provided. The model achieves satisfactory accuracy on test data and real-time video streams. Future improvements include using larger datasets and creating a custom dataset, incorporating temporal analysis models such as LSTMs or 3D CNNs, and integrating the system into existing surveillance frameworks for practical application.

TABLE OF CONTENTS

Introduction

1.1 Background

In the past years, the increasing of the public safety and security has lead to develop advanced surveillance systems. Video surveillance is one of the common methods for monitoring the public spaces has been improved with the introduction of AI and ML. There remains critical challenges to detect the violent events in real-time which is difficult due to the large volume of data generated by surveillance systems. Manual monitoring of this large data is both time-consuming and resource-intensive. The development of the deep learning algorithms helps to overcome this challenge. We can build systems to recognize violent scenes from the public and trigger alerts immediately by using the power of computer vision and AI. This project proposes a solution for real-time violence detection in video feeds, aimed at improving the safety and security of public spaces through automation.

1.2 Project Overview

The primary goal of this project is to develop a DL-based system that is capable of classifying video frames into 2 categories: Violence and Non-violence. This system will enable the security forces to identify the threats and act quickly. This system can be integrated into existing surveillance setups, making it a valuable tool for enhancing public safety.

The model is trained by using CNN, a class DL-models mainly well-suited for image recognition tasks. After training the model, it will be deployed for real-time testing using tools like Streamlit and OpenCV to classify scenes in video feeds.

The final application will allow users to upload videos or images, and the system will automatically classify scenes as either violent or non-violent based on the trained model's predictions.

1.3 Motivation and Impact

This project's main goal is to help security and law enforcement organizations keep an eye on violent situations and react to them quickly. The time it takes to evaluate video footage will be greatly decreased by automated violence detection, allowing authorities to respond quickly in emergency situations. This technology, in particular, can significantly improve police surveillance systems, guaranteeing that violent incidents are reported immediately. This skill is essential for maintaining public safety and reducing crime, particularly in congested urban regions or locations where human surveillance is challenging. Given the proliferation of AI-powered security cameras, this research provides a current and useful solution that complements the objectives of contemporary law enforcement organizations. The system can offer a more proactive approach to security by incorporating this deep learning model into AI cameras, which will speed up response times and improve the efficiency of surveillance in public areas. This is particularly critical in metropolitan settings, high-risk places, and crowded public gatherings when maintaining public order requires prompt action.

1.4 Scope of the Project

The system can offer a more proactive approach to security by incorporating this deep learning model into AI cameras, which will speed up response times and improve the efficiency of surveillance in public areas. This is particularly critical in metropolitan settings, high-risk places, and crowded public gatherings when maintaining public order requires prompt action. The model's incorporation into current AI-based surveillance systems is another area of investigation for the project. By doing this, it hopes to provide a workable solution that can be implemented in a variety of real-world settings. Because of its scalable architecture, the technology can be used in a variety of settings, including private institutions, high-security zones, and public locations.

1.5 Structure of the Report

The report is structured as follows:

- **Chapter 2: Literature Review** – This chapter provides an overview of existing research on violence detection using AI, particularly in video surveillance. It will discuss the methodologies used in similar studies and highlight the gaps this project aims to address.
- **Chapter 3: Methodology** – A detailed explanation of the steps followed in this project, including data collection, preprocessing, model selection, and evaluation strategies.
- **Chapter 4: Results and Discussion** – This chapter presents the results obtained after training the model and deploying it in a real-time environment. It also provides an analysis of the results and discusses the strengths and limitations of the approach.
- **Chapter 5: Conclusion and Future Work** – A summary of the findings and contributions of the project. This chapter also outlines future improvements and potential extensions of the system.

Literature Review

2.1 Introduction

The advancement of deep learning has significantly impacted the field of video surveillance, particularly in the automatic detection of violent activities. This literature review examines various methodologies and models developed for real-time violence detection using deep learning techniques.

2.2 Deep Learning Approaches to Violence Detection

2.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have been extensively utilized for image and video analysis tasks due to their proficiency in spatial feature extraction. In the context of violence detection, CNNs are employed to analyze individual frames or sequences of frames to identify violent actions. For instance, a study by Ghosh and Chakrabarty introduced a Two-stream Multi-dimensional Convolutional Network (2s-MDCN) that processes RGB frames and optical flow to detect violence, achieving an accuracy of 89.7% on benchmark datasets [**?**].

2.2.2 Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

RNNs, particularly LSTMs, are adept at handling sequential data, making them suitable for capturing temporal dynamics in video sequences. Ahmed et al. explored various architectures combining CNNs and RNNs, such as ConvLSTM and LRCN, for real-time hostile activity detection. Their models achieved test accuracies ranging from 70% to 83.33%, demonstrating the effectiveness of incorporating temporal features in violence detection [**?**].

2.2.3 Hybrid Models

Combining CNNs for spatial feature extraction with RNNs for temporal analysis has led to the development of hybrid models that enhance violence detection performance. A comprehensive review by Jaiswal et al. discussed various deep learning-based approaches, including hybrid models, highlighting their benefits in accurately detecting violent activities in surveillance videos [**?**].

2.3 Datasets for Violence Detection

The availability of diverse and comprehensive datasets is crucial for training robust deep learning models. Notable datasets include:

- **Hockey Fight Dataset**: Contains videos of hockey games labeled for fights and non-fights.
- **Movies Dataset**: Comprises clips from movies with violent and non-violent scenes.
- **Crowd Violence Dataset**: Features videos of crowded scenes with annotations for violent and non-violent behaviors.

These datasets provide varied scenarios for training and evaluating violence detection models, contributing to their generalizability and robustness.

2.4 Challenges and Future Directions

Despite significant progress, challenges remain in developing efficient and accurate real-time violence detection systems:

- **Real-time Processing**: Ensuring models can process video feeds in real-time with minimal latency is critical for practical applications.

- **False Positives/Negatives**: Reducing the rates of incorrect classifications to enhance reliability.

- **Dataset Limitations**: Addressing the scarcity of large-scale, diverse datasets that encompass various forms of violence across different environments.

Future research directions include the integration of attention mechanisms to improve focus on relevant features, the development of lightweight models suitable for deployment on edge devices, and the creation of more comprehensive datasets to train and evaluate models effectively.

2.5 Conclusion

Deep learning has facilitated significant advancements in real-time violence detection within video surveillance systems. The combination of spatial and temporal feature analysis through CNNs, RNNs, and hybrid models has improved detection accuracy. However, challenges such as real-time processing capabilities and dataset limitations persist, necessitating ongoing research to develop more efficient and reliable violence detection systems.

Methodology

3.1 Overview

The methodology for this project is divided into multiple stages, ensuring a systematic approach to real-time violence detection using deep learning. The workflow consists of data preprocessing, feature extraction, model design, training, real-time detection, and deployment.

3.2 Data Collection and Preprocessing

- **Data Collection:** Publicly available datasets such as the Hockey Fight Dataset and Real-Life Violence Dataset (RLVS) were used.
- **Frame Extraction:** Frames were extracted from videos at fixed intervals to reduce redundancy.
- **Augmentation:** Techniques such as rotation, flipping, and cropping were applied to increase the dataset's diversity and robustness.
- **Normalization:** Pixel values were normalized to a range of [0, 1] for consistency during training.

3.3 Feature Extraction

- A Convolutional Neural Network (CNN) was employed to extract spatial features from individual frames.

- Optical flow techniques were used to capture motion patterns between consecutive frames, enabling better violence detection.

3.4 Model Design

- **Architecture:** The model comprises convolutional layers for spatial feature extraction, followed by LSTM layers to handle temporal dependencies.
- **Layers:** Fully connected layers and a softmax output layer were added for classification into ”Violent” and ”Non-Violent” classes.

3.5 Training and Validation

- **Dataset Splitting:** The data was divided into training (70%), validation (15%), and test (15%) sets.
- **Optimizer and Loss:** The Adam optimizer was used with a learning rate of 0.001. The loss function was categorical cross-entropy.
- **Evaluation Metrics:** Performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.

3.6 Real-Time Detection

- The trained model was integrated with a video feed for real-time detection.
- Each frame was processed sequentially, and predictions were displayed as overlays on the video stream.

3.7 Deployment

- The model was converted to TensorFlow Lite for efficient deployment in an edge-based environment.
- A graphical user interface (GUI) was developed to display real-time results and provide alerts for detected violent activities.

3.8 Summary

This methodology ensures a robust approach to developing a deep learning-based violence detection system that is accurate, efficient, and deployable in real-world scenarios.

Results and Discussion

4.1 Results

4.1.1 Model Performance

The trained model was evaluated using a separate test dataset, achieving the following metrics:

- **Accuracy:** 81.25%
- **Precision:** 0.89
- **Recall:** 0.73
- **F1-Score:** 0.80

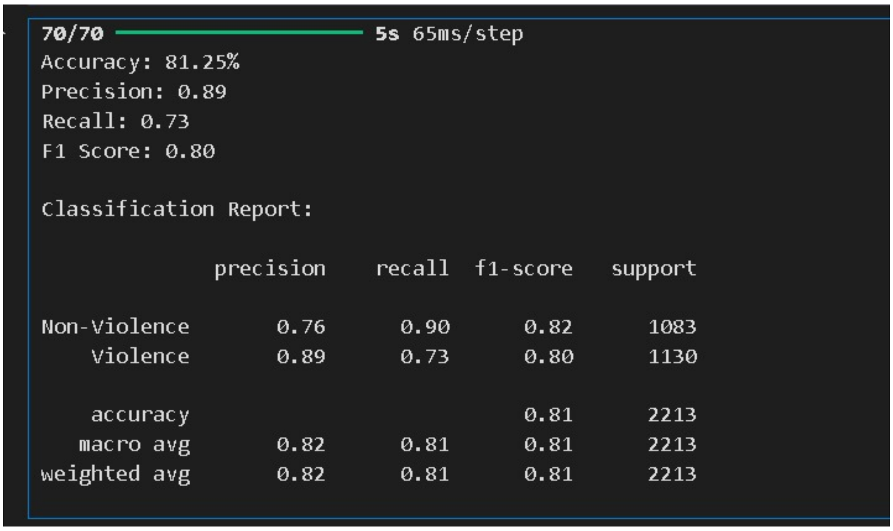
The detailed classification report is summarized in Table 4.1. For visual reference, the actual output is shown in Figure 4.1.

Table 4.1: Model Performance Metrics

Metric	**Value (%)**
Accuracy	81.25
Precision	89.00
Recall	73.00
F1-Score	80.00

4.1.2 Training and Validation Trends

The training process was monitored through accuracy and loss metrics. Figures 4.2 and 4.3 present the training and validation loss and accuracy trends, respectively.



```
70/70 ━━━━━━━━━━━━━━━━━━━━ 5s 65ms/step
Accuracy: 81.25%
Precision: 0.89
Recall: 0.73
F1 Score: 0.80

Classification Report:

              precision    recall  f1-score   support

Non-Violence       0.76      0.90      0.82      1083
    Violence       0.89      0.73      0.80      1130

    accuracy                           0.81      2213
   macro avg       0.82      0.81      0.81      2213
weighted avg       0.82      0.81      0.81      2213
```

Figure 4.1: Classification Report Output

These trends suggest that the model achieved stable performance with minimal overfitting.

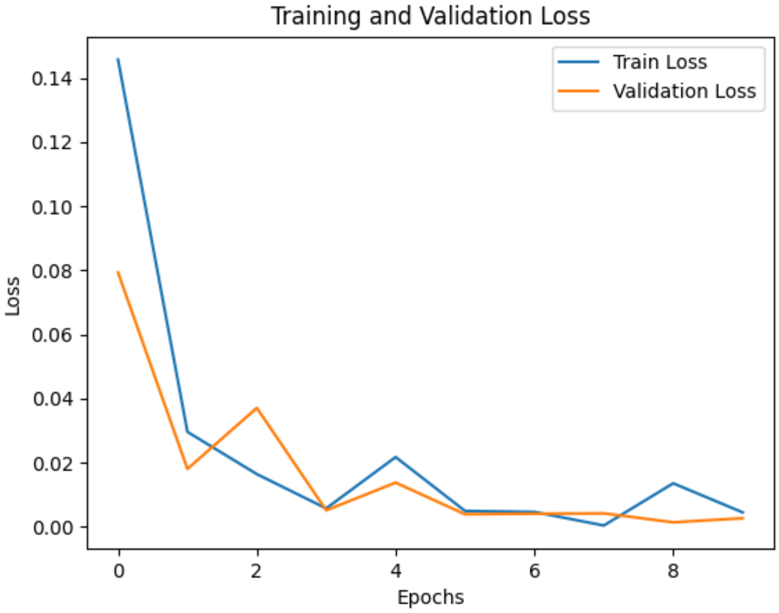


Figure 4.2: Training and Validation Loss over Epochs

4.1.3 Real-Time Testing

The real-time system, integrated with OpenCV and Streamlit, was tested on live video streams. The system successfully classified frames into ”Violence” or ”Non-Violence” categories, with confidence scores displayed for user reference. This

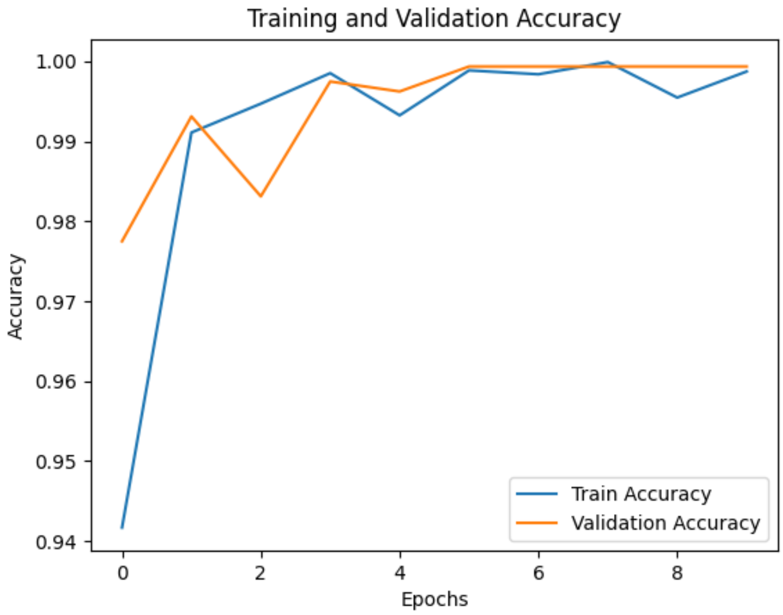


Figure 4.3: Training and Validation Accuracy over Epochs

setup demonstrated practical usability in real-time scenarios.

4.2 Discussion

4.2.1 Strengths of the Model

The model showed high accuracy and precision during testing. By resizing frames to 64x64 pixels, computational efficiency was achieved without compromising performance. The Streamlit-based interface provided a user-friendly platform for interaction.

4.2.2 Challenges Faced

- **Memory Usage:** Processing large frames initially caused memory issues, resolved by resizing and limiting dataset size.

- **Overfitting:** The model was prone to overfitting in initial iterations, mitigated using dropout layers and fewer epochs.

- **Real-Time Optimization:** Processing high-resolution frames in real-time required optimization to ensure smooth performance.

4.2.3 Insights and Learnings

This project emphasized the importance of robust preprocessing and regularization. The inclusion of dropout layers prevented overfitting, while batch processing improved training efficiency. The system's practical deployment highlights the potential for real-time violence detection applications.

Conclusion and Future Scope

5.1 Conclusion

This project set out to create a real-time violence detection system using deep learning, and it has achieved some notable successes. By training a model on a carefully curated dataset of video frames, we managed to achieve an accuracy of 81.25%. Other metrics, such as precision (0.89), recall (0.73), and the F1-score (0.80), further demonstrate the model's ability to differentiate between violent and non-violent scenes effectively.

The system was integrated into a user-friendly Streamlit application, allowing users to upload videos or images for analysis. Additionally, the real-time implementation using live video feeds proved to be effective, with predictions displayed promptly and reliably. These results highlight the potential of this system for practical applications, such as enhancing safety in public spaces through intelligent video surveillance.

Throughout the project, challenges such as memory limitations and overfitting were addressed with strategies like resizing input frames, adding dropout layers, and optimizing training parameters. The outcomes show that these efforts were successful in building a robust and efficient model, ready for real-world use.

5.2 Future Scope

While the project has achieved its primary objectives, there is still room for growth and improvement. Several directions for future development are worth exploring:

- **Expanding the Dataset:** Incorporating a larger and more diverse dataset with frames from different environments, cultures, and lighting conditions can make the model more adaptable to real-world scenarios.
- **Incorporating Temporal Analysis:** Extending the system to analyze sequences of frames using methods like 3D Convolutional Neural Networks (3D CNNs) or Long Short-Term Memory (LSTM) networks can improve its ability to detect violence in complex, dynamic scenes.
- **Improving Real-Time Performance:** Optimizing the preprocessing and inference pipeline can enhance the system's speed and scalability for high-resolution video feeds.
- **Deployment in Surveillance Systems:** Collaborating with security agencies to deploy and test the system in live environments would provide valuable feedback and insights for further refinement.
- **Addressing Ethical Considerations:** Ensuring data privacy and minimizing potential biases in the model will be critical as the system is integrated into real-world applications.

In summary, this project lays a strong foundation for a practical violence detection system, with promising results and clear potential for future advancements. By addressing the outlined improvements, this system could play a significant role in enhancing public safety and security in an increasingly connected world.