



Violence Detection Using Deep Learning

Krishna Sapagale¹, Manoj Sanikam², Nikitha³, Prajwal M Shetty⁴, **Kiran B V⁵**

¹ Alva's Institute of Engineering and Technology, Mijar, krishnasapagalev130@gmail.com

² Alva's Institute of Engineering and Technology, Mijar, manojksanikam01@gmail.com

³ Alva's Institute of Engineering and Technology, Mijar, nikithaswamy@gmail.com

⁴ Alva's Institute of Engineering and Technology, Mijar, prajwalshetty63615@gmail.com

⁵ Alva's Institute of Engineering and Technology, Mijar, kiranbv@aiet.org.in

Received Date : November 23, 2023 Accepted Date : December 22, 2023 Published Date : January 07, 2024

ABSTRACT

Due to the increased risk of exposure to violent and harmful content brought about by the spread of online video content, robust systems for automatic detection and filtering have to be developed. This research suggests a novel method for deep learning-based violent content detection in videos. Our model examines both temporal and spatial characteristics in video frames by utilizing the power of recurrent neural networks (RNNs) and convolutional neural networks (CNNs). The suggested system uses a two-stream architecture, where one stream is used for temporal information using bidirectional LSTM (Long Short-Term Memory) networks to capture sequential dependencies, and the other stream is devoted to spatial analysis using 3D CNNs for frame-level understanding [1]. To ensure strong generalization, the model is additionally trained on a varied dataset that includes both violent and non-violent content. Transfer learning is used with pre-trained deep learning models on large-scale datasets to improve the model's performance [5]. Comprehensive tests show how well the suggested method works to reliably identify violent content in videos of different genres and settings. The system demonstrates its potential for incorporation into online video platforms to give viewers a safer and more secure experience by achieving state-of-the-art outcomes in terms of precision, recall, and F1 score [4]. The suggested deep learning-based approach supports further initiatives to lessen the negative impacts of violent content in digital media and promote a safe and healthy online community [1]. Using Deep Learning to Address the Problem of Violent Video Detection: A Bright Future for Security and Safety.

The proliferation of violent content is a key concern posed by the ever-increasing abundance of online video content. This puts personal safety, public safety, and platforms' capacity to properly filter information at risk. Presenting deep learning, a potent technique that presents a viable way to automatically identify violent content in videos [2]. To sum up, deep learning presents a potent and exciting way to address the pressing problem of violent video content. We can create a more secure online environment for everyone by utilizing this technology properly and resolving the issues it raises [5]. Further investigation into cross-modality learning and real-time detection shows promise for even higher efficiency and accuracy.

Key words: Deep Learning Methods, Multi Model Feature Extraction, Machine Learning, Fight, Violent Flow, Motion feature extraction, Feature fusion baseline.

1. INTRODUCTION

Due to the ongoing rise in abnormal behaviour in different contexts, human behaviours detection in general and violence detection in particular have recently gained significant attention in Computer Vision (CV) research. Additionally, because of the complexity of the environment (i.e., social interaction) and the challenge associated with extracting a particular characteristic that is associated with a particular occurrence, violence detection is one of the most challenging problems in CV [3].

To put it another way, accurately detecting a violent situation requires two main feature extraction methods: 1) Spatial or shape feature extraction, and 2) Temporal or time features extractions. The spatial features represent the relationships or interactions between single frame pixels, but they are insufficient to identify the violence.

In the meanwhile, the most well-liked study in violence detection uses surveillance footage to extract spatiotemporal elements that aid in the clear identification of violent cases. In order to improve overall classification performance, this paper proposed various architectures based on extracting spatiotemporal features using various techniques (e.g., 3D Convolutional Neural Network (CNN) Convolutional Long Short-Term Memory (Conv-LSTM) networks integrating transfer learning with LSTM or Conv-LSTM). Additionally, the architectures included a combination of attention modules (i.e., channel attention and spatial attention).

Based on the UBI-Fights video data, a great deal of important work has been done recently in the area of violence detection. For instance, in order to provide weak/self-supervised learning, Bruno Manuel Degradin in suggests a complex iterative learning framework based on Bayesian filtration for the instances of unlabeled input. Further more, the author employed the late decision fusion ensemble technique to improve the overall performance of three models using the random forest algorithm, which has fifty decision trees [2].

The results showed that this framework performs 0.819 for the Area Under the Curve (AUC) metric and 0.284 for the Equal Error Rate (EER) measure on the UBI-fights data. Proposing different architectures based on integrating the Convolutional Block Attention Modules (CBAM) with various layers such as ConvLSTM2D or Conv2d&LSTM layers; to catch the spatiotemporal features, and increase the focus on the important ones.