**KianNet: A Violence Detection Model Using an Attention-Based CNN-LSTM Structure**

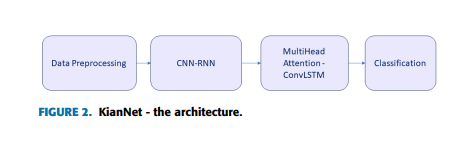
**INTRODUCTION:**

The increasing demand for comprehensive public safety monitoring through video surveillance has led to challenges in detecting abnormal behaviors, such as violence, due to the vast amounts of data and limited labeled anomalies. To address this, a novel approach combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) has been proposed. This method utilizes CNNs to extract spatiotemporal features from individual frames and RNNs, specifically Convolutional LSTM (ConvLSTM), to analyze temporal relationships among these features. The paper introduces a unique model, KianNet, which merges Multi-Head Self-Attention (MHSA) and ResNet50-ConvLSTM architectures to improve violence identification by capturing complex spatiotemporal features. Evaluation on datasets like UCF-Crime and RWF demonstrates the model's superiority over existing algorithms. The paper aims to address the urgent need for effective violence detection systems in public safety surveillance, contributing to enhanced public safety and security.

**METHODOLOGY:**

1. 3D-CNN: Utilizes convolutional layers and pooling layers to extract spatiotemporal features from video sequences, achieving state-of-the-art performance in violence detection tasks by leveraging spatial and temporal information.
2. CNN-RNN: Combines the strengths of Convolutional Neural Networks (CNNs) for spatial feature extraction and Recurrent Neural Networks (RNNs) for capturing temporal dependencies in video sequences, resulting in improved performance in violence detection by effectively capturing spatial and temporal features.
3. Attention-Based Models: Utilize selective focus on specific parts of input data, allowing neural networks to weigh the importance of different features selectively. By selectively attending to informative frames or regions within a video, attention-based models enhance performance in violence detection tasks, achieving higher accuracy while reducing computational cost.

Top of Form



**SUMMARY OF FINDINGS:**

The paper introduces KianNet, a violence detection approach for surveillance footage, combining ResNet50 for feature extraction, ConvLSTM for frame sequence analysis, and MHSA for vision saccade emulation. Experimental results on UCF-Crime and RWF datasets showcase KianNet's superior performance in binary classification for violence detection. Despite its effectiveness, KianNet's high parameter count prompts future work on a lightweight attention mechanism. Additional improvements include integrating YOLOv3 for action recognition, expanding KianNet's application to areas like fall detection, and enhancing feature extraction with original images and sound inputs, highlighting KianNet's potential in diverse domains beyond violence detection, such as healthcare.

**FUTURE SCOPE:**  
In future endeavors, the focus will be on refining the KianNet approach to violence detection. Efforts will include reducing the number of training parameters by developing a lightweight violence detection attention mechanism. Furthermore, there's a plan to explore the integration of YOLOv3 for recognizing actions post-feature extraction, potentially enhancing video content understanding. Additionally, expanding the application of KianNet to domains like fall detection in healthcare settings is envisioned, leveraging its robust learning structure. Moreover, improvements in feature extraction will be pursued by integrating original images with moving parts from frame subtractions. Lastly, enhancing violent action detection accuracy by incorporating sound inputs is proposed, acknowledging the significance of audio cues in comprehensive video analysis. These future directions aim to advance the efficacy and versatility of KianNet for violence detection and broader video analysis applications.