ABSTRACT

In a modern civilized community, public safety is of prime importance, and the detection of anomalous events has become a vital factor for a successful security system. Conventional Video Surveillance (VS) methods are inadequate for identifying anomalous events by themselves. It is due to their inability to analysis large sequential data in dynamic environments. Video Anomaly Detection (VAD) has undergone swift development with the emerging Artificial Intelligence (AI) technologies. The research addresses the challenges of conventional VAD and proposes hybrid Deep learning (DL) models using transfer learning techniques for an efficient VAD system in dynamic environments.

The application range of VAD is not limited to but includes social, commercial and industrial surveillance systems. Traffic management in urban areas, crowd management, emergency response and resource optimization are VAD's key application fields. The wide requirement for intelligent VAD inspires the design and development of reliable and efficient video surveillance systems capable of automating Anomaly Detection (AD) with minimal human intervention.

The study develops and evaluates four hybrid models for VAD using Deep Learning techniques based on transfer learning to obtain improved performance. The first research phase proposed a CNN-YOLO hybrid model capable of anomaly detection. This model uses CNN for model training and modified YOLOv4 for object detection, ensuring accurate and high-speed anomaly detection. This model processes a single random frame out of 100 input frames and yields a faster response.

The CNN-YOLO have high accuracy and faster response but being a small model, it samples a random frame input only. To overcome this limitation and the inability of sequential video processing of the CNN-YOLO model, a hybrid model comprised of Residual Network (ResNet) and Long Short-Term Memory (LSTM) was executed in the second phase. This model can execute feature extraction and sequential information processing in more than thousands of video frames. ResNet-50 is employed for spatial feature extraction and LSTM to capture temporal relationships of the input video data even though this hybrid model enhances detection capability but has low accuracy, efficiency and generalization skill due to overfitting.

In the third research phase, a segmentation-based anomaly detection technique is implemented, reducing overfitting. This model is constituted by hybridizing Improved UNet (IUNet) with the Cascade Sliding Window Technique (CSWT). In the IUNet-CSWT hybrid model, standard convolutional layers of IUNet are replaced with a ConvLSTM for spatiotemporal feature extraction. CSWT estimates the anomaly score of the input video and thus classifies it to normal and anomalous events. This model is equipped for processing complex patterns since it has an effective equilibrium between generalization skill and precision. A low false positive rate and high detection accuracy make the model work effectively in crowded environments.

The fourth phase implemented a Hierarchical Multiscale-CNN with LSTM model, enhancing multi-scale feature identification and temporal data analysis. This model can work efficiently even in low-resolution video utilizing a Bilateral-Wave Denoise Technique. Multi-scale CNN is augmented by a Spatial Pyramid Pooling (SPP), which enhances the feature extraction. The performance of this model outperforms all the other models.