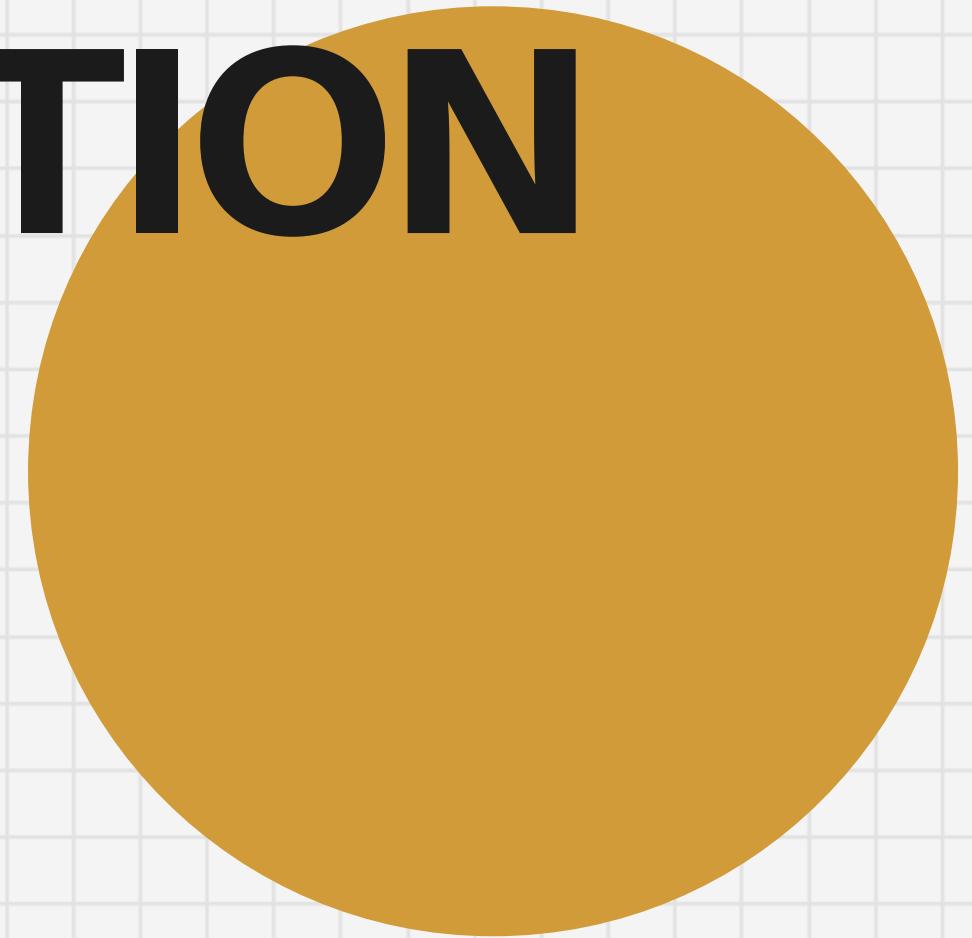
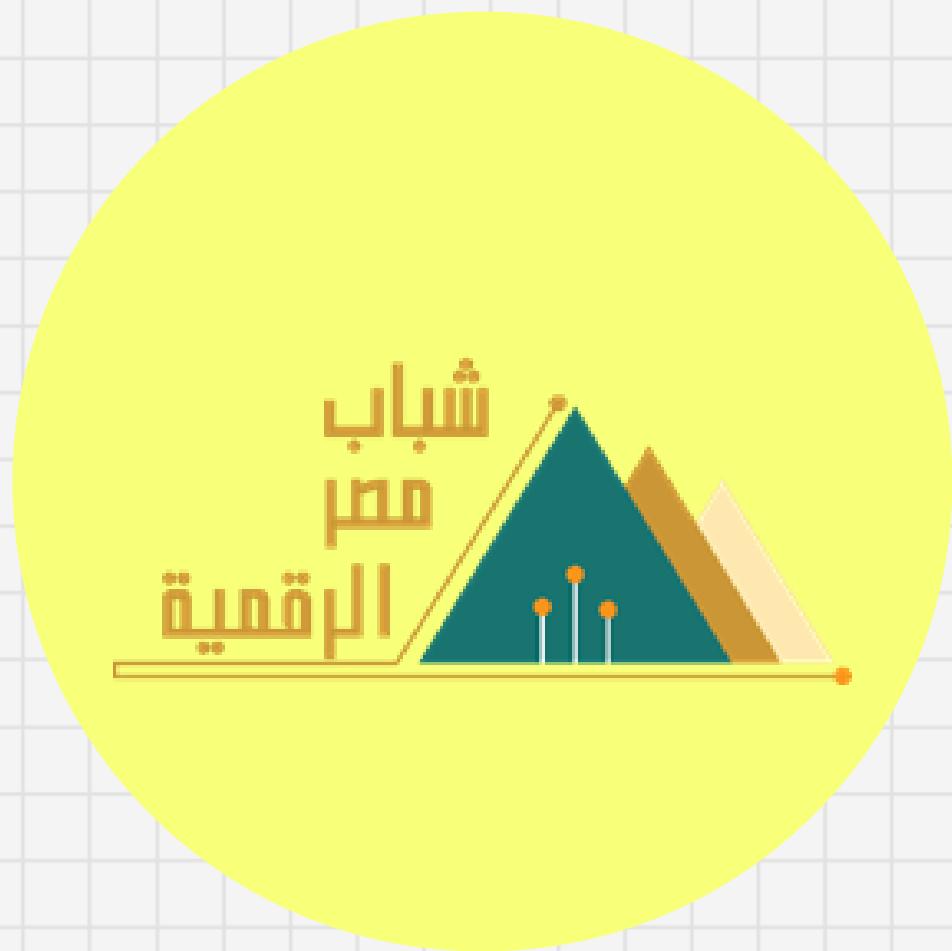


NATIONAL TELECOMMUNICATION INSTITUTE

REAL TIME VIOLENCE DETECTION

AI and IoT DEY Initiative



Introduction



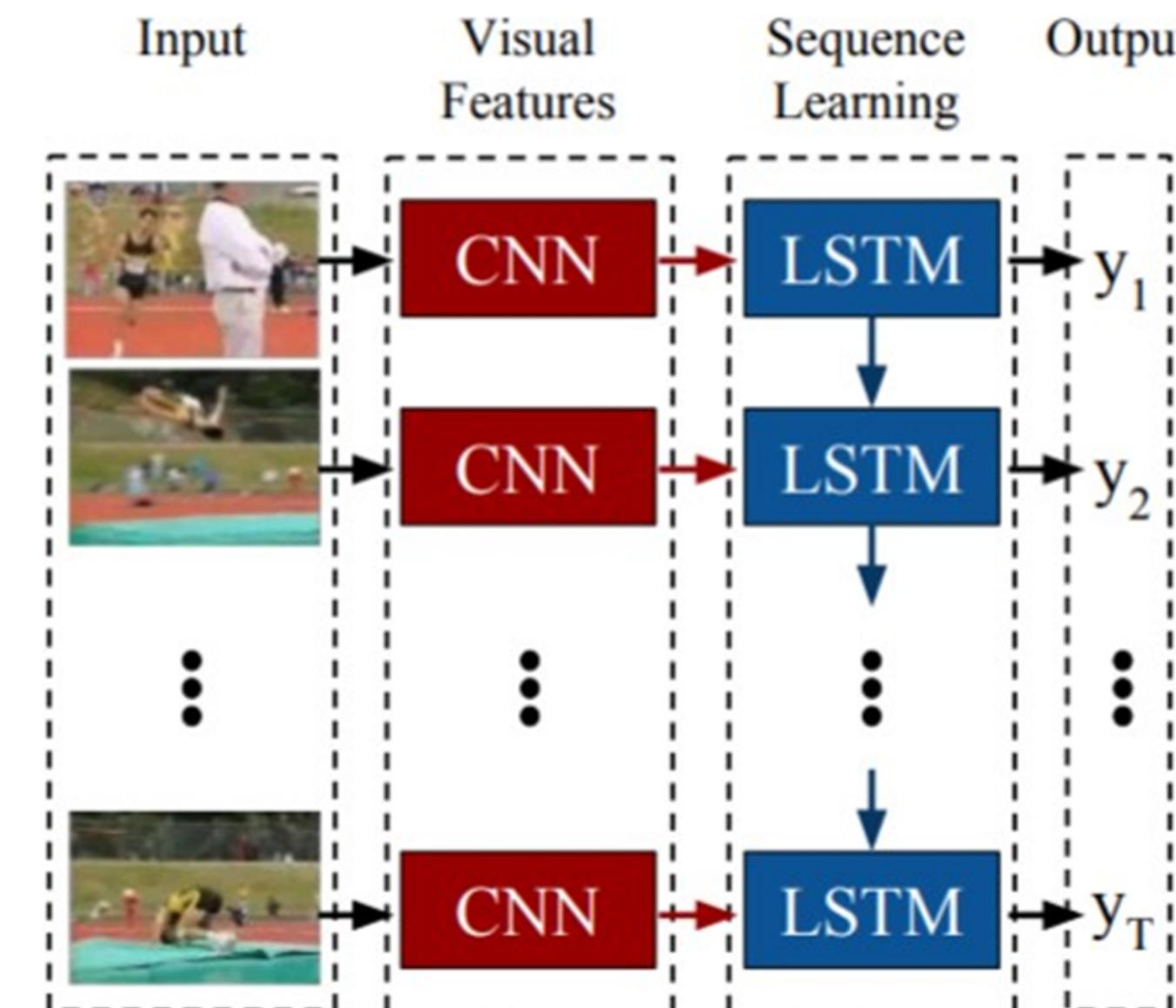
Detection of a violence event in surveillance systems is playing a significant role in law enforcement and city safety. The effectiveness of violence event detectors measures by the speed of response and the accuracy and the generality over different kind of video sources with a different format. Several studies worked on the violence detection with focus either on speed or accuracy or both but not taking into account the generality over different kind of video sources.



Abstract

In this Work, we proposed a real-time violence detector based on deep-learning methods. The proposed model consists of a MobileNet Pretrained Model as a spatial feature extractor and Bidirectional LSTM as temporal relation learning method with a focus on the three-factor (overall generality - accuracy - fast response time). The suggested model achieved about 97% accuracy with speed of 16 frames/sec.

The dataset contains only two directories: Non-Violence (which contains 1000 real life situations videos like eating, sports activity, singing, etc. And the other directory Violence (contains 1000 videos with severe violence in various situations).



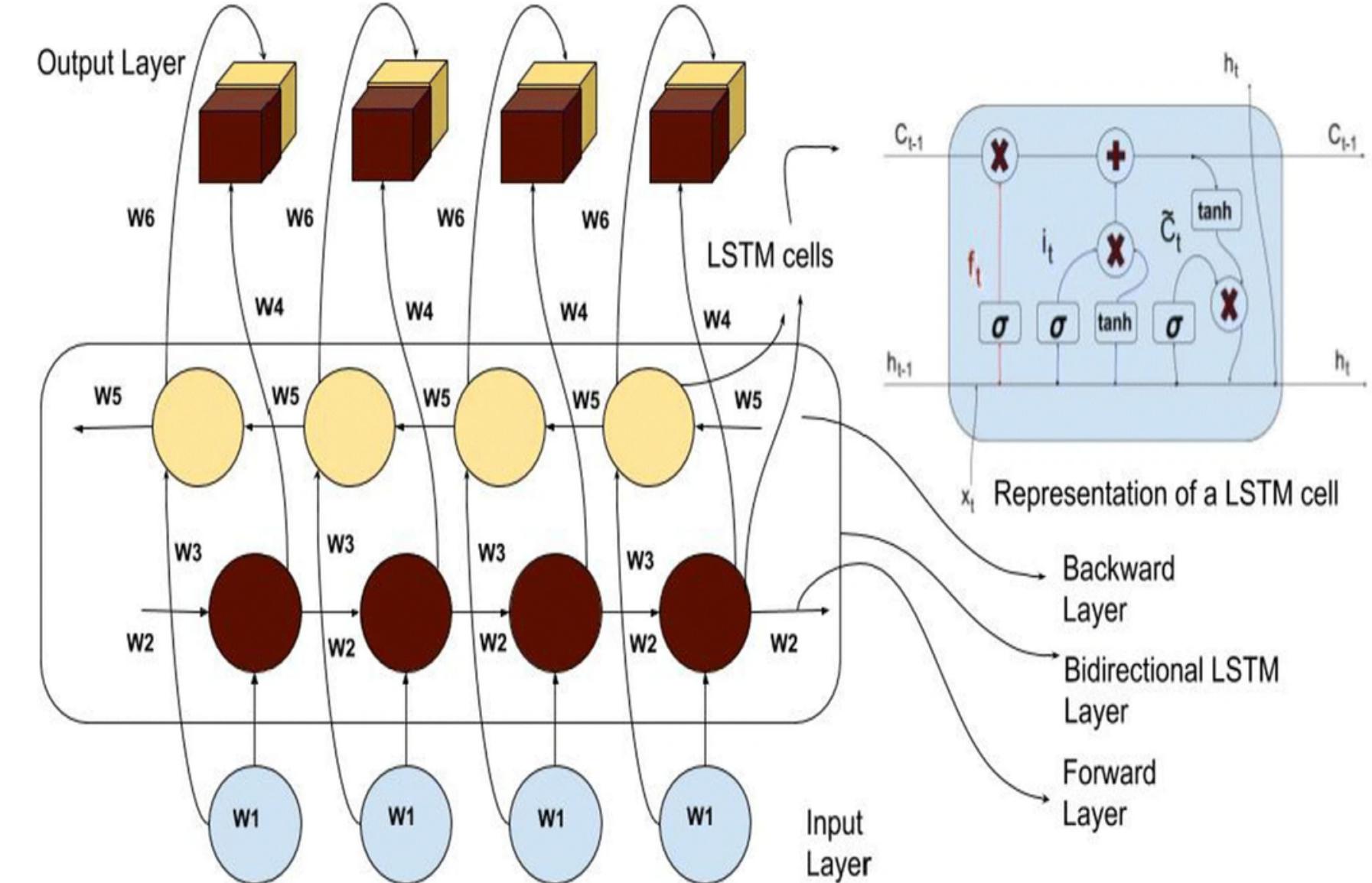
Hypothesis

To classify violent or non-violent actions, our model must be able to predict

sequences in consecutive frames, that is a pattern in the movement of the subjects or a degree of their motion, etc.

This is not possible by considering only the spatial features (features belongs to a particular frame) of the frames.

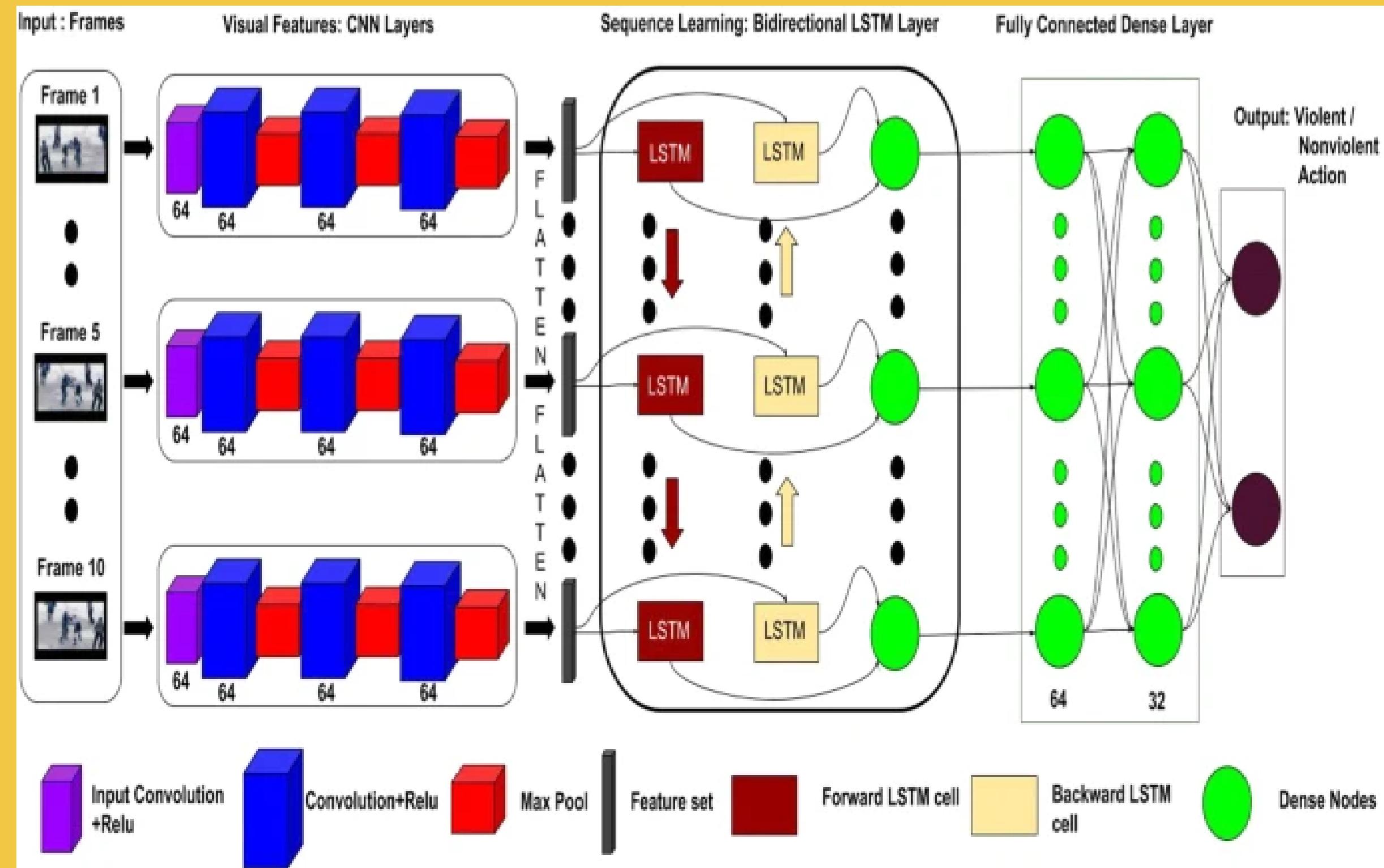
The temporal or time-related features must also be considered while detecting sequences in the frames. The temporal features may be processed in the direction of upcoming frames or reverse order.



Methodology

Our model processes the temporal features in both the direction in addition to the spatial features, which helps the model to become more accurate at the same time consumes less computational time. The lightweight models are always preferred in surveillance due to its low-cost structure.

- MobileNet as a Feature Extractor
- Bidirectional LSTM
- Dense Layers

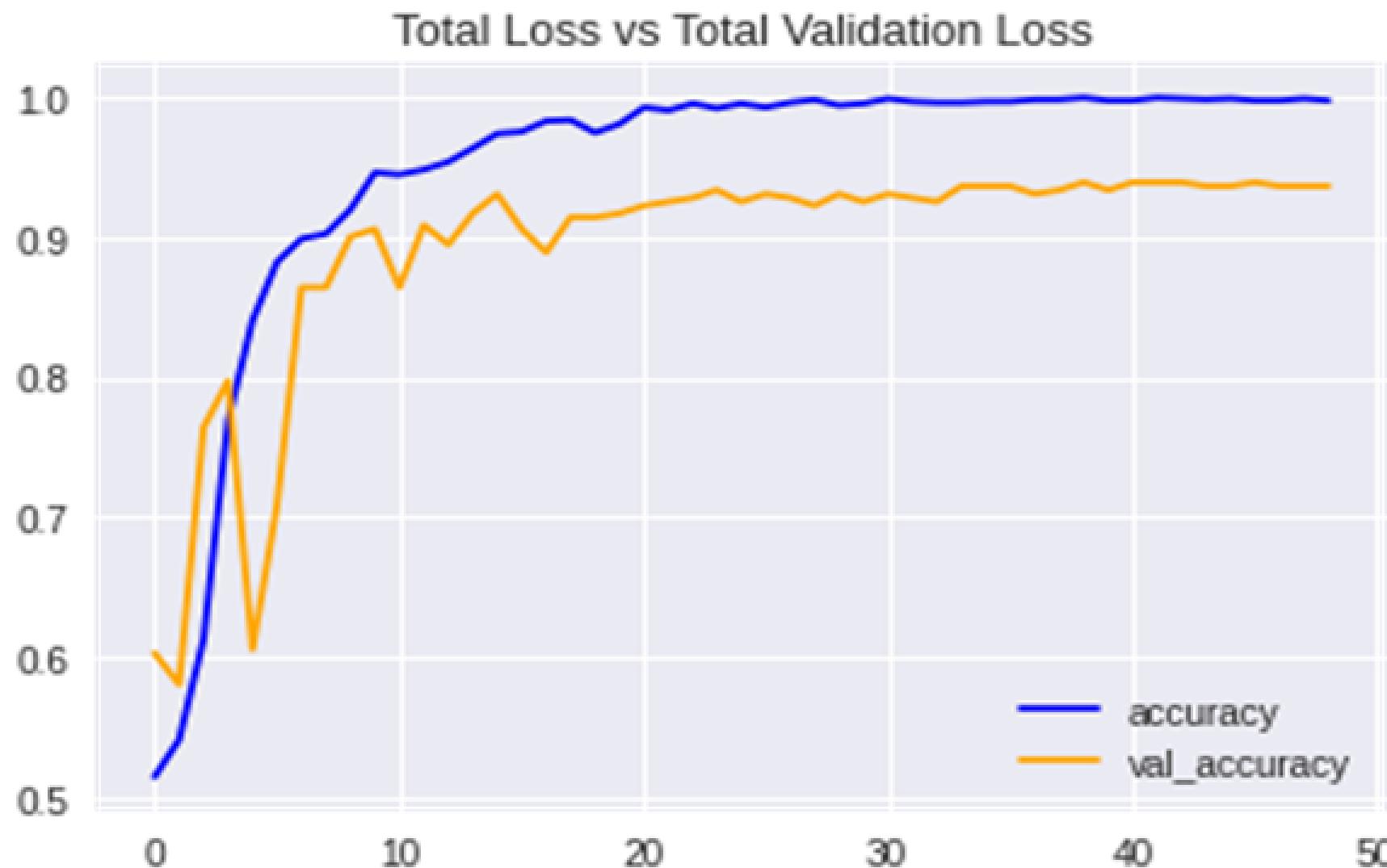


Results

◆ We attained to achieve significant results with our proposed model.

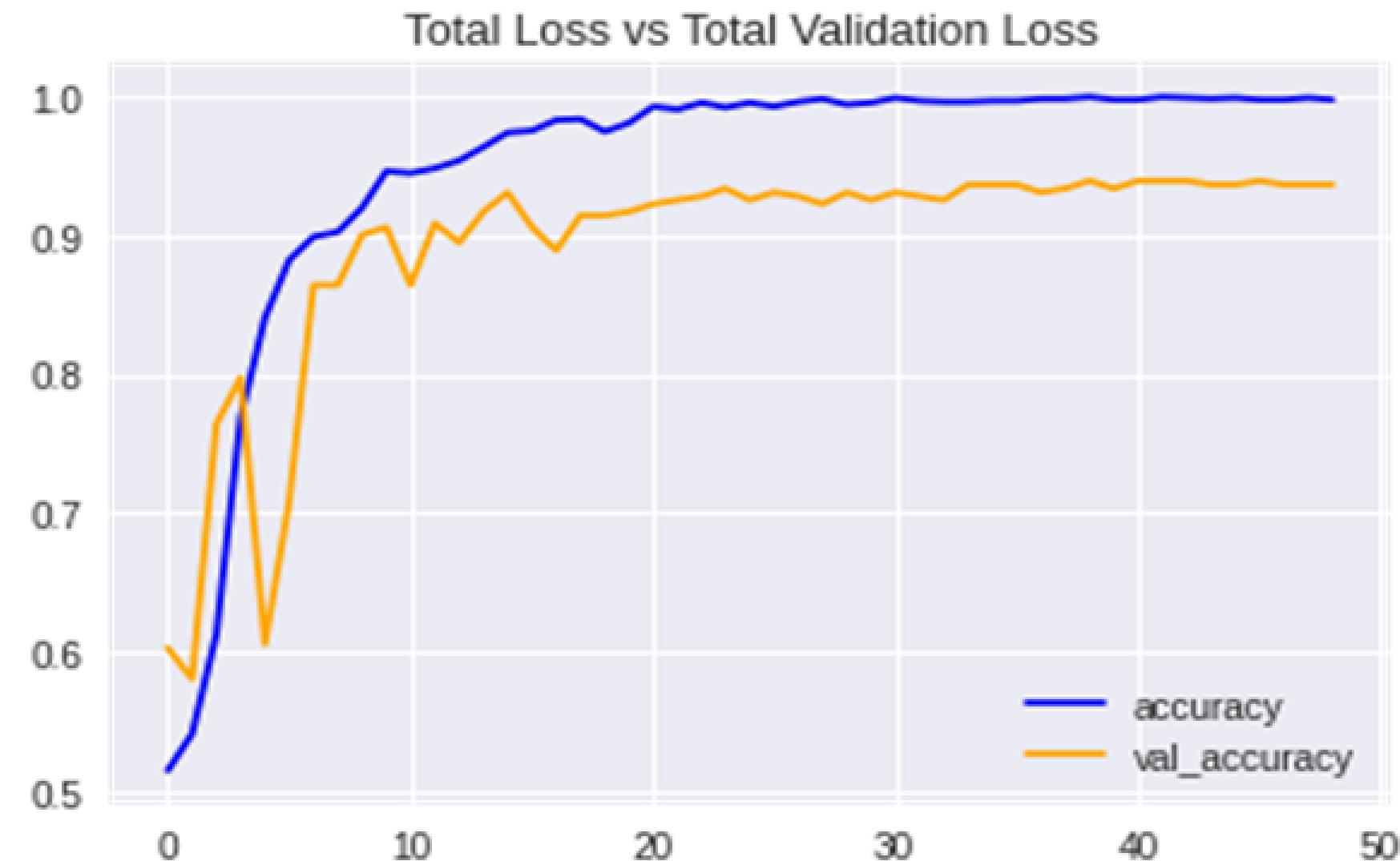
- Accuracy

Our Model achieved about 94–97%
Accuracy On the Testing Set.

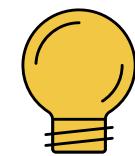


- Loss

Our Model achieved about 0.15–0.25
Loss On the Testing Set.



Expirments



We Implemented Two Prediction Functions To Test Our Model On.

- First Function Perform Frame By Frame Prediction For The Video.



- Second Function Perform Prediction For The Whole Video.

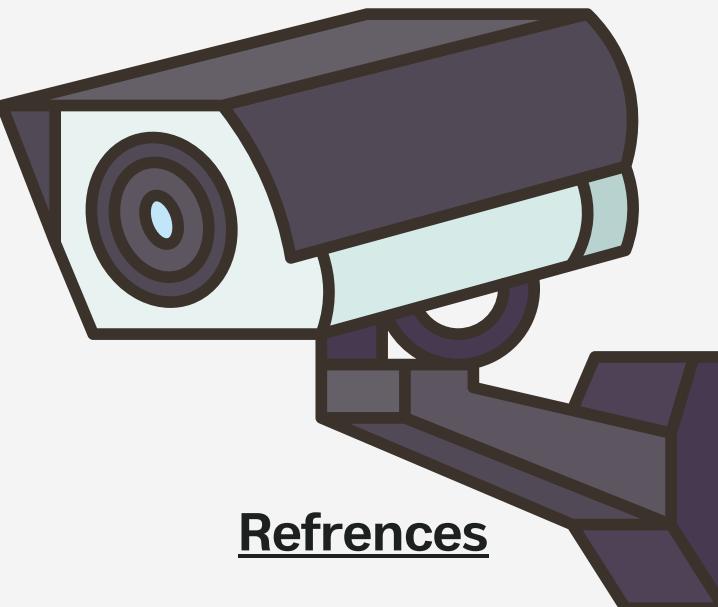
Predicted: Violence
Confidence: 0.9999997615814209



Summary and Conclusion



- Our proposed MobileNetV2-BiLSTM variant provides the reportedly best results for the used dataset.
- Despite the performance of our proposed model, it needs to be further validated with more standard datasets where identification of one to many or many to many violent activities including weapons are tough to detect.
- the Detection model can be extended and be deployed on an IOT device like a security cam or a surveillance cam , So the device can notify the system admin or the responsible authorities if there's any Violence behavior detected



References