Video-Based Violence Detection System

Presented by:
Thatipamula Soumya - 100521729084
Namdhari Anusha - 100521729082

Problem Statement

Traditional surveillance systems face challenges with real-time and accurate detection of violent activities, often relying on limited automated systems or manual review of video feeds, both of which can be slow. This project aims to develop an automated violence detection system that processes video footage in real-time using deep learning models. By combining MobileNetV2 for spatial feature extraction and LSTM for temporal analysis, the system can accurately classify activities as "Normal" or "Violence." To ensure timely intervention, a Telegram bot sends instant alerts with relevant details and evidence to authorities.

Dataset

SmartCity CCTV Violence Detection Dataset (SCVD) is having 2000 videos
Dataset structure:

- Train and Test folders with three classes:
    - Normal, Violence, and Weaponized.
- the "Weaponized" class is merged with the "Violence" category to simplify the classification problem and enhance the model's focus
- Sample videos per category visualized for understanding.

Data Preprocessing

- **Frame Extraction**: A fixed number of frames (15) are extracted from each video to maintain temporal consistency.
- **Frame Resizing**: Resizing each frame to a uniform size of 128x128 pixels.
- **Normalization**: Scaling pixel values to [0,1] for model compatibility.
- **Class Label Encoding:** Merging Violence and Weaponized into a single label, Violence, for binary classification.
- **One-Hot Encoding**: Labels are converted to one-hot encoded vectors to suit the neural network training.

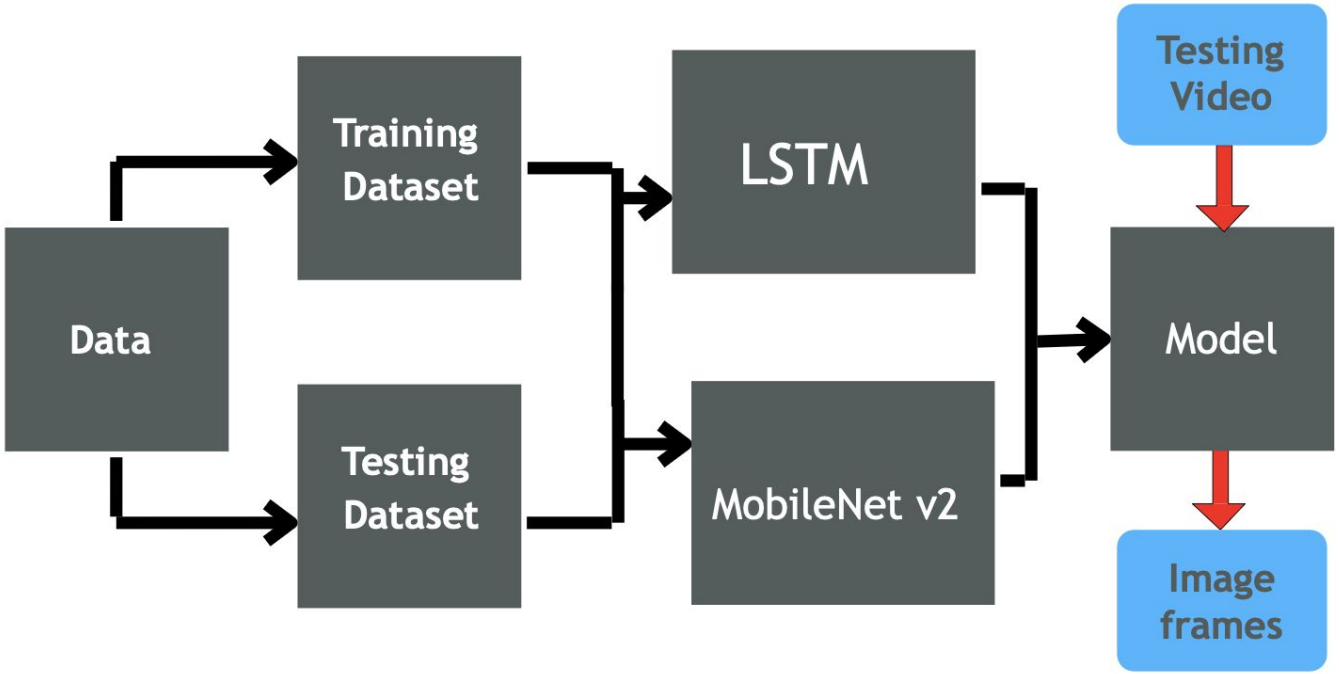
MobileNetV2

- MobileNetV2 is used for extracting spatial features from video frames.
- Pre-trained on ImageNet, this model leverages depth-wise separable convolutions to achieve high efficiency with minimal computational cost.
- Its lightweight architecture makes it suitable for real-time applications.
- MobileNetV2 is fine-tuned for the specific task of violence detection.
- By freezing its layers during training, the model retains its pre-trained knowledge, reducing the computational burden and training time.
- The extracted features are then passed through the LSTM layers for temporal analysis, forming a hybrid model capable of capturing both spatial and temporal dynamics.

LSTM

- LSTM layers are for analyzing the sequential nature of video data.
- After spatial features are extracted by MobileNetV2, they are fed into stacked LSTM layers to model temporal dependencies across frames.
- The LSTM architecture enables the system to recognize patterns over time, such as the progression of violent activities.
- Two LSTM layers, configured with 128 and 64 units respectively, are used to enhance the temporal feature extraction capability.
- This architecture allows the model to identify subtle temporal cues that distinguish violent actions from normal behavior, significantly improving classification accuracy.

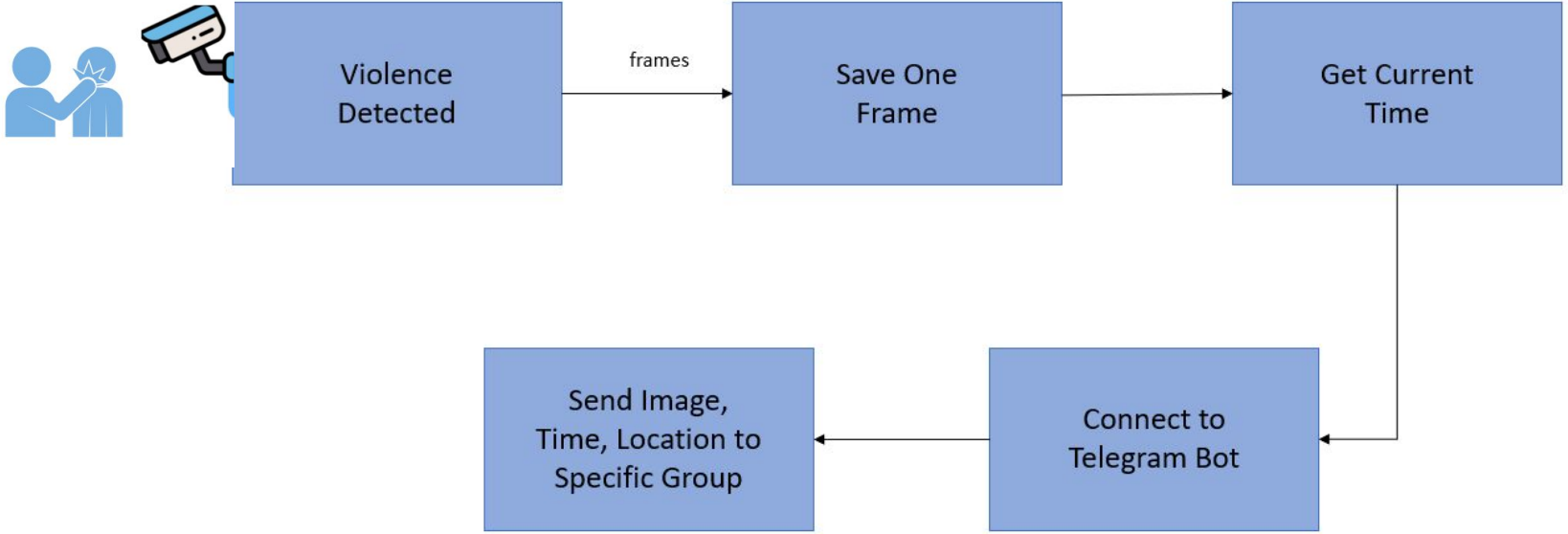
Architecture Diagram



Telegram Bot

- When violence is detected in a video, the bot sends real-time alerts to a specified group or individual, including the incident's timestamp, location, and a snapshot of the detected event.
- This functionality is implemented using the Telepot library, which enables seamless communication between the deep learning model and Telegram.
- The bot ensures timely responses to critical situations by delivering instant notifications, bridging the gap between detection and action.

Architecture Diagram



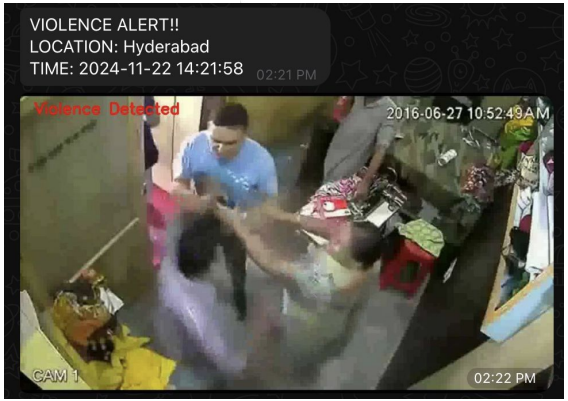
Training and Validation

- Model built using LSTM layers on extracted frame features for sequence analysis.
- Hyperparameters:
  - Optimizer: Adam with a learning rate of 1e-3.
  - Loss function: Binary Cross Entropy (for binary classification).
  - Validation split: 20%.
- Early Stopping and Model Checkpointing to monitor validation loss, prevent overfitting, and save the best model.

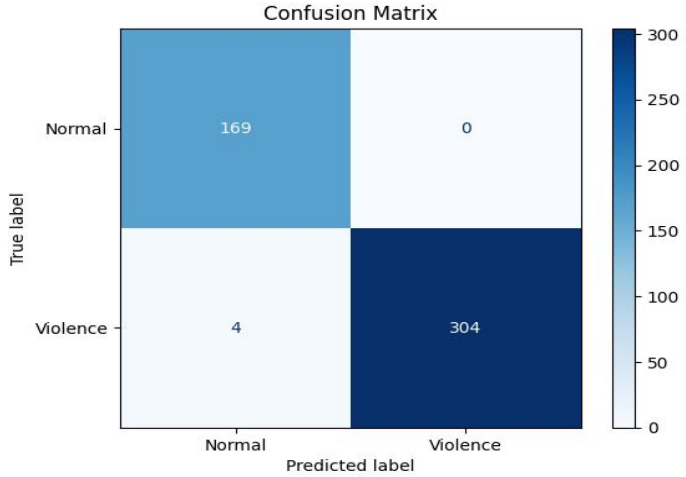
Results


Normal
2012-11-09 18:35:5
IPC


Violence Detected
23-04-2018 11 07 01 AM
Facebook / Carnes Cares
DOS CAJAS

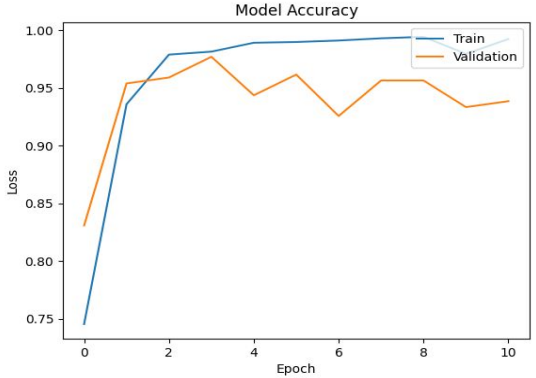

VIOLENCE ALERT!!
LOCATION: Hyderabad
TIME: 2024-11-22 14:21:58
02:21 PM
Violence Detected
2016-06-27 10:52:49 AM
CAM 1
02:22 PM

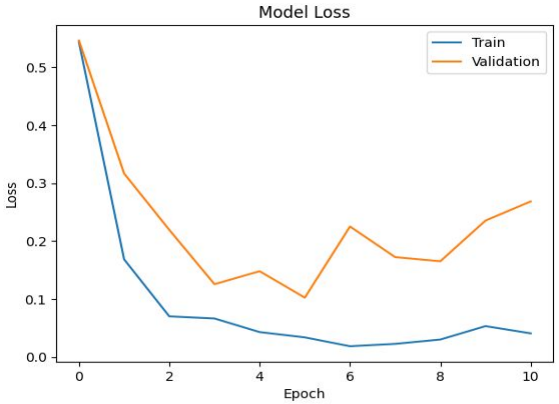
Evaluation



15/15 ━━━━━━━━━━━━━━━━━━━━ **47s** 3s/step

	precision	recall	f1-score	support
0	0.98	1.00	0.99	169
1	1.00	0.99	0.99	308
accuracy			0.99	477
macro avg	0.99	0.99	0.99	477
weighted avg	0.99	0.99	0.99	477





Conclusion

- This project presents a robust solution for real-time violence detection using a combination of MobileNetV2 and LSTM architectures.
- The integration of a Telegram bot ensures timely communication of critical alerts, enhancing the system's applicability in surveillance settings.
- By preprocessing video data effectively, the system achieves high accuracy and generalization.
- **Future enhancements** could include training on a larger dataset, incorporating additional features such as object detection, and optimizing the system for deployment on edge devices.

References

1. SmartCity CCTV Violence Detection Dataset (SCVD) from kaggle
2. Aremu, T., Zhiyuan, L., Alameeri, R., Khan, M., & Saddik, A. E. (2024). "SSIVD-Net: A Novel Salient Super Image Classification and Detection Technique for Weaponized Violence." In K. Arai (Ed.), Intelligent Computing (pp. 16–35). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-62269-4_2
3. TensorFlow: https://www.tensorflow.org/
4. Keras API: https://keras.io/
5. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
6. Telepot Library: https://telepot.readthedocs.io/en/latest/
7. Python's datetime and pytz Libraries: https://docs.python.org/3/library/datetime.html
8. OpenCV Library: https://opencv.org/