Real-time Crime Detection Using Customized CNN

Samundeswari S
Department of Computer Science and Engineering,
Sri Sairam Engineering College,.
Chennai, India
samundeswari.cse@sairam.edu.in

Harini M
Department of Computer Science and Engineering,
Sri Sairam Engineering College,.
Chennai, India
sec19cs016@sairamtap.edu.in

Dharshini T
Department of Computer Science and Engineering,
Sri Sairam Engineering College,.
Chennai, India
sec19cs021@sairamtap.edu.in

Srinithi S
Department of Computer Science and Engineering,
Sri Sairam Engineering College,.
Chennai, India
sec19cs031@sairamtap.edu.in

2022 1st International Conference on Computational Science and Technology (ICCST) | 978-1-6654-7655-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCST55948.2022.10040379

***Abstract*— Every day, there are more crimes committed and criminals are on the loose, which is making people fear for their safety. The primary goal is to detect and deter illicit activity before it occurs. With the aid of cutting-edge technology, CCTV is commonly used in both private and public spaces. It is possible to control crime in this area, but human supervision is required to oversee it. It's difficult for a human to keep track of multiple screens at the same time. Human error is a possibility in many situations. To overcome this drawback, we stipulate a Deep Learning-based Real-Time Crime Detection Technique that analyzes real-time CCTV footage and alerts a nearby supervisor about the crime in the current region. The model tracks the movement of people and classifies it as aggressive or nonviolent behavior using the Multiple Object Detection with Localization technique. Any aggressive conduct filmed by the camera will be detected and instantaneously alerted by the system.**

Keywords—Deep learning, Object tracking, Crime detection , CNN, alert system

I. INTRODUCTION

The rate of recorded criminal occurrences every day is significantly rising. Robbery, vandalism, assault, murder, kidnapping, etc. are all examples of crimes. Criminal investigation and crime prevention are both crucial, but they are frequently carried out manually, which is quite ineffective. The surveillance monitoring system that is currently in use requires a lot of manual labor. The procedure of manual observation and information collecting from video footage takes more time.

We propose a DL-based crime detection system called "SpotCrime" to increase public safety and aid law enforcement in reducing the amount of human labor required when identifying crimes. SpotCrime is a web application that monitors live CCTV footage and alerts the police officials if any suspicious activity is found.

The system automates the process of monitoring video surveillance by assessing human activity based on behavior and notifying the officials through the SpotCrime application if any violent or suspicious activity is found. With feature extraction at the frame level, the behavior classification process begins. Then, utilizing Customized CNN and cutting-edge models, the final pooling layer output is collected from the video frames. In order to recognize suspicious activity, CNN models were built from the ground up with a variety of convolutional and dense layer counts. Following comparisons of the validation accuracy of each model for the application, the best model was selected.

Need for the system – Crime has increased in India, particularly during lockdowns. As the number of criminal activities has increased, it has become critical to protect people's safety to the maximum extent possible. To promote public safety, police officers' main priority is to prevent crime from occurring. Circuit television cameras (CCTVs) are employed for surveillance to keep an eye on local crime. However, regular human monitoring is required, which is complex and prone to human error. We presented a Real-Time Crime Detection System that uses deep learning to watch real-time videos and alert police personnel of a crime's incidence and location.

II. RELATED WORK

A related work to our model is [1]. The authors have suggested a technique to monitor criminals in live video feeds, spot firearms in public places that are unnecessary, and contact the local police cybercrime admin in order to improve public safety and reduce crime. A Crime Detection System based on the YOLO algorithm was suggested, and with a final average loss of 0.6 and a mean average precision of 78.3%, it can successfully identify criminals even in congested areas.

Another work which explained more about the YOLO algorithm was [2]. Yolov5 is a potential architecture for real-time facial detection. In terms of real-time facial recognition systems, the authors compared the most recent YOLOv5 with earlier iterations (v3 and v4). Experiments reveal that the newly produced version YOLOv5 is 87 percent accurate on Face Detection Data Set and Benchmark and 94 percent accurate on a real-time face recognition tailored dataset, which is significantly better than prior versions. It is self-evident that deep learning-based facial recognition will outperform all other disciplines of study.

The authors in [3] demonstrated a system for enhancing public safety and assisting law enforcement that was based on machine learning and deep learning. It combines two key elements: a crime predictor based on machine learning and an automated surveillance notification system based on deep learning. Additionally, the monitoring component of video surveillance includes identifiers for suspicious people and odd activity. By assessing human activity based on behavior and informing the appropriate parties through a mobile application of any aggressive or suspicious activities, the system automates the process of video surveillance monitoring. Frame-level feature extraction is the first step in the behavior classification process.

The authors in [4] had proposed the Crime Intent Detection System, an automated system that prevents crimes by identifying a pistol or knife in a person's hands using a VGGNET19 pre-trained model that computes more quickly than the GoogleNet Inception V3 model. When someone is seen with weapons like a gun or knife, this system transmits the crime intention detection security message to the registration number. In comparison to other methods now in use for crime detection, this approach yields good results. To detect crime scenes such as people holding a gun or knife in their hands, the device might be integrated into television. The TV can transmit automated crime intent security alerts to a registered number in the event that a crime is committed.

According to the authors in [5], numerous research have revealed some nonverbal cues that, when combined with our intellectual discernment, can help us successfully detect a person with dubious motives. Nonverbal cues including covering the mouth, stroking the neck, and keeping hands in a pocket are chosen as suspicious behaviors. Convolution Neural Network (CNN) is the architecture employed for non-verbal gesture training and classification. The trained model is then implemented for real-time inference in an image recognition programme. The system uses numerous cameras and the Jetson TX1 development kit from NVIDIA® to instantly identify suspicious non-verbal cues.

An alert is issued to the user if more than two nonverbal cues from the same topic are found.

III. PROPOSED WORK

A. *Dataset Collection*

This model makes use of the Kaggle dataset for Real-Life Violence Situations. 1000 violent and 1000 non-violent videos are included in this dataset. This dataset of violent videos includes numerous instances of actual street brawls occurring in a variety of settings. The non-violence videos in our collection are drawn from a variety of human activities, including sports, eating, walking, etc. Additionally, YouTube video site scraping was carried out in order to create a model with adequate accuracy. Online scraping was carried out using the Selenium Python module, a potent tool for web automation and programmatic control of a web browser. For getting videos from the internet, Pytube, a lightweight, dependency-free Python package, was utilized. These downloaded videos are saved in the same location where the software is run. Fig 1(a). shows the data collected from Kaggle and Fig. 1(b). shows the data collected from web scraping.

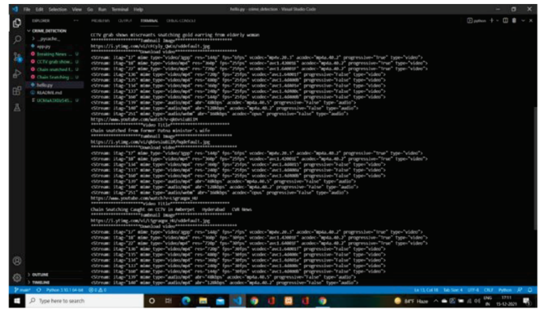

Fig. 1(a). Collected data from Kaggle.

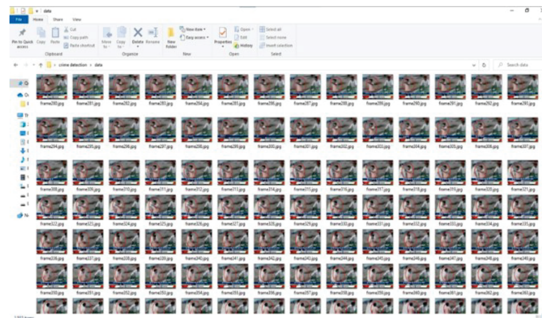

Fig. 1(b). Collected data from web scraping.

B. *Data Pre Processing*

Both the violent and non-violent videos are extracted, then they are transformed into frames and kept in ../Data/Training/V and ../Data/Training/NV at a frame size of 150x150 and a frame skip of 2. Data preprocessing entails the augmentation of data through a variety of random modifications, and ensures no image is seen by the model repeatedly. This keeps the model from overfitting and improves generalization.

This is accomplished using the class in keras : keras.preprocessing.image.ImageDataGenerator. Data generators are inputs to Keras model techniques.

- The rotation range, which provides a range within which images will be rotated at random, is a number in degrees (0–180).
- Within the width shift and height shift constraints, images can be arbitrarily translated vertically or horizontally (measured as a percentage of the overall width or height).
- The input will be multiplied by a factor called "rescale" before any more processing is done. Our original photographs' RGB values vary from 0 to 255, but at a typical learning rate, these values would be too high for our models to handle. In order to obtain values between 0 and 1, we therefore scale our original photos by a ratio of 1/255.
- Shearing transformations can be applied at random with shear range.
- Zoom range allows for arbitrary image zooming
- When there is no assumption of horizontal asymmetry, the horizontal flip command randomly flips half of the images horizontally

- The Fill mode technique is used to fill in newly formed pixels, which may emerge following a rotation or a width/height shift.

C. Training Model Using CNN

The pre-processed dataset is trained by using deep learning algorithm CNN and using Keras. The CNN algorithm can take in an input image and differentiate different elements and objects in an image by assigning them different weights and biases. 32 filters are used. Then, max pooling is applied to minimize the spatial dimensions of the output volume. 64 filters are then applied. Again max pooling is used to minimize the spatial dimensions. Finally 128 filters. The customized CNN model with optimized parameters performed well on the training and sufficiently good in the test dataset. The architecture is currently being used by the website and test.ipynb file.

D. Evaluation using the trained model and creating an alert system

The trained model is used to detect if there is any suspicious activity found in live feed from CCTV cameras. The live feed from a CCTV camera is split into frames using OpenCV and it is classified as suspicious or non-suspicious using the trained model. The consecutive frames threshold is taken as 20 so that if more than 20 consecutive frames have been marked suspicious, the system will ring an alarm and alert the officials. A website is designed using the Django framework in such a way that we have to choose a video file to test the model. The trained model is used to identify if the action in the uploaded video is suspicious or non-suspicious. If it is found suspicious, an alert sound is produced and a warning sign is shown on the screen.

IV. RESULTS

The system proposes real-time weapon and criminal face detection. Since the development of CNN, real-time object detection has advanced significantly, which gave rise to numerous object detection methods such as YOLO, SSD, and others. When it comes to running the detection of criminals and weapons, CNN is pretty effective. The fastest performer is CNN, which processes one image in 44 milliseconds.

The model's accuracy during training was 90.7 percent, and the loss was 23.88 percent. The model's validation loss is 66.80 percent, while validation accuracy is 76.05 percent. Adam is the optimizer employed to improve the model. The epoch value of the model is 30. This technology makes it possible to track criminals in live video feeds, find unnecessary weapons in public areas, and alert the local authorities about cybercrime, improving public safety and reducing crime.

Fig 3(a). shows the accuracy curve and Fig 3(b). shows the loss curve for the Customized CNN model during the training and validation process.

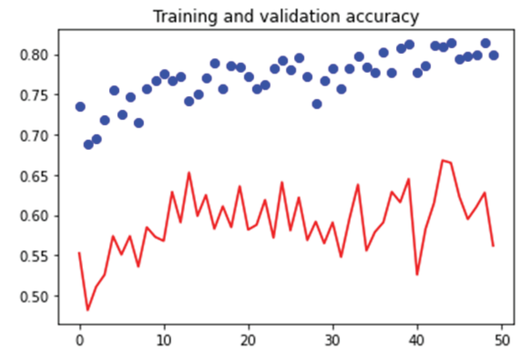

Fig. 3(a). Accuracy curve for Customized CNN model.

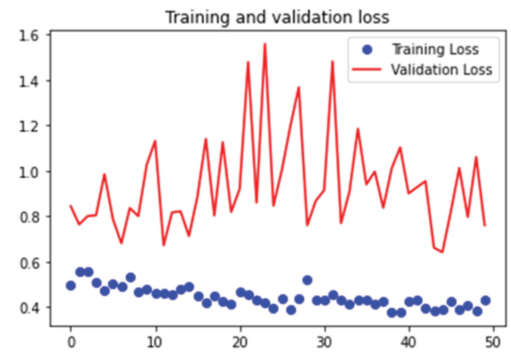

Fig. 3(b). Loss curve for Customized CNN model.

V. CONCLUSION

In order to promote societal security and lower crime rates, this research study offers a method to identify criminal activity in real-time and notify police officers. With a training accuracy of 90.7% and a loss on training of 0.4292, we suggested a real-time criminal detection system employing the Customized CNN algorithm that can successfully identify criminals even in crowded areas.

REFERENCES

[1] P. Sivakumar, J. V, R. R and K. S, "Real Time Crime Detection Using Deep Learning Algorithm," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021.

[2] F. Majeed, F. Z. Khan, M. J. Iqbal and M. Nazir, "Real-Time Surveillance System based on Facial Recognition using YOLOv5," 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), 2021.

[3] C. Rajapakshe, S. Balasooriya, H. Dayarathna, N. Ranaweera, N. Walgampaya and N. Pemadasa, "Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction," 2019 International Conference on Advancements in Computing (ICAC), 2019.

[4] U. V. Navalgund and P. K., "Crime Intention Detection System Using Deep Learning," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018.

[5] F. T. George, V. S. P. Patnam and K. George, "Real-time deep learning based system to detect suspicious non-verbal gestures," 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2018.

[6] W. Safat, S. Asghar and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 9.J.

[7] A. Abhari, S. Banihashemi and J. Li, "Object Movement Detection by Real-Time Deep Learning for Security Surveillance Camera," 2017

International Conference on Computational Science and Computational Intelligence (CSCI), 2017.

[8] M. Nieto, L. Varona, O. Senderos, P. Leskovsky and J. Garcıa, "Real-time video analytics for petty crime detection," 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), 2016

[9] Zhang, Da & Maei, Hamid & Wang, Xin & Wang, Yuan-Fang. (2017). Deep Reinforcement Learning for Visual Object Tracking in Videos.

[10] Nayak, S., Ravindran, B. (2020). Reinforcement Learning for Improving Object Detection. In: Bartoli, A., Fusiello, A. (eds) Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science(), vol 12539. Springer.

[11] H. Nam and B. Han. Learning multi-domain convolutional neural networks for visual tracking. arXiv preprint arXiv:1510.07945, 2015.

[12] Uzkent, Burak & Yeh, Christopher & Ermon, Stefano. (2019). Efficient Object Detection in Large Images Using Deep Reinforcement Learning

[13] Piyush Chhoriya, "Automated Criminal Identification System using Face Detection and Recognition", International Research Journal of Engineering and Technology, October 2019.

[14] Pratibha, Akanksha Gahalot, Uprant, Suraina Dhiman, Lokesh Chouhan, "Crime Prediction and Analysis", Institute of Electrical and Electronics Engineers, Stanford University, 2017.