iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** IV **Month of publication:** April 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.59968

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue IV Apr 2024- Available at www.ijraset.com*

# Real-Time Violence Detection in Surveillance Videos Using Deep Learning Approach

Prof. N.P. Mohod[1], Aadesh Sawarkar[2], Vrushali Sirsath[3], Shrutika Tayde[4], Sanjana Bhagat[5], Hitesh Surve[6]
[1]*Assistant Professor,* [2, 3, 4, 5, 6]*Final Year, Department of Computer Science & Engineering, Sipna College of Engineering and Technology, Amravati, Maharashtra, India.*

***Abstract:*** ***In order to tackle the ongoing problem of keeping fruits fresh, this research presents an automated freshness detection. The ever-growing presence of surveillance cameras has undoubtedly enhanced public safety. However, manually reviewing vast quantities of video footage for violence detection remains a tedious and error-prone task. This paper proposes a deep learning-based approach for anomaly detection in surveillance videos, with a specific focus on identifying violent activities. Traditional surveillance methods often lack the sophistication to distinguish between normal behaviour and potentially threatening actions. Deep learning offers a significant advantage by automating the identification of deviations from established behavioural patterns in video data. This automation enables real-time analysis of footage, potentially signifying the occurrence of violent incidents and allowing for a swifter response from security personnel. This research delves into the application of deep learning models for violence detection in surveillance footage. We begin by discussing the limitations inherent in conventional methods, such as motion detection and manual review, which struggle to capture the nuances of human behaviour. Subsequently, we explore the advantages that deep learning-based approaches bring to the table, including their ability to learn complex patterns from large datasets and identify subtle anomalies that might escape human observation. The methodology section details the training process for these models and investigates suitable deep learning architectures specifically tailored for violence detection tasks. Finally, the paper concludes by presenting the anticipated performance improvements achievable with this approach, such as increased accuracy and reduced false alarms. We also outline potential future research directions in this evolving field, such as improving interpretability of deep learning models and addressing privacy concerns associated with video surveillance.***

***Keywords: Surveillance video, Anomaly detection, Deep learning, Convolutional Neural Networks (CNNs), MobileNetV2, Security***

## I. INTRODUCTION

The widespread adoption of surveillance cameras has transformed security landscapes. These systems act as a powerful deterrent against criminal activity, providing valuable evidence for investigations and fostering a sense of security in public spaces. Traditionally, security personnel have relied on manual monitoring of video feeds, a tedious and error-prone process. This approach struggles with the sheer volume of footage generated, leading to missed incidents and inefficient resource allocation. Imagine a security guard tasked with watching dozens of screens simultaneously – fatigue, distraction, and information overload become inevitable. Recent advancements in artificial intelligence, particularly deep learning, offer a revolutionary solution for video surveillance: anomaly detection. This technology automates the identification of deviations from established behavioural patterns in video footage. By analyzing video data and learning normal activity, anomaly detection systems can flag unusual occurrences that might signify potential security threats. Imagine a system that can continuously monitor hundreds of cameras, intelligently filtering out normal activity and alerting security personnel only when something suspicious happens. This allows security personnel to focus their attention on these flagged events, improving responsiveness and overall security effectiveness. The integration of deep learning into anomaly detection offers significant advantages over traditional methods. Deep learning models excel at extracting complex features from video data, enabling them to identify subtle anomalies that might escape human observation. For instance, a deep learning model could detect a person loitering in a restricted area, even if the person is simply standing still. Additionally, these models can continuously learn and improve over time, adapting to changing environments and evolving security threats. Imagine a system that can learn from past security breaches, becoming increasingly adept at identifying suspicious behaviour. This continuous learning capability makes deep learning-based anomaly detection a powerful tool for proactive security management, allowing security personnel to stay ahead of potential threats. However, the implementation of deep learning-based anomaly detection also presents challenges that need to be addressed. One key concern is the vast amount of data required to train these models effectively.

Additionally, ensuring the privacy of individuals captured on video footage is paramount. Balancing security needs with privacy concerns requires careful consideration and the implementation of appropriate safeguards. Finally, the interpretability of deep learning models remains an ongoing area of research. Understanding how these models reach their decisions is crucial for building trust in their effectiveness. Despite these challenges, deep learning-based anomaly detection holds immense promise for revolutionizing video surveillance, creating a future where intelligent systems augment human security efforts and contribute to safer communities.

This paper investigates the potential of the MobileNet V2 model for anomaly detection in surveillance videos. We evaluate its performance using standard metrics like accuracy, loss, and F1-score, along with ROC curves and AUC. By analyzing these metrics, we aim to assess the model's effectiveness in distinguishing normal activities from anomalies and contribute valuable insights into the practical application of deep learning for video surveillance. Furthermore, we explore techniques to address potential limitations, such as data efficiency and interpretability. This research contributes to the ongoing development of robust and reliable deep learning models for anomaly detection, paving the way for a future where intelligent video surveillance systems enhance security and safety in our communities.

## II. LITERATURE REVIEW

C B Murthy et. al proposed a system to detect object in images/videos using various deep learning techniques and embedded platforms such as convolutional neural network (CNN), regions with convolutional neural networks (R-CNN), spatial pyramid pooling layer (SPPNet), you only look once, version 3 (YOLOv3), Fast-RCNN, computer vision(CV), graphics processing unit (GPU). Different dataset used here for training and testing the model of 330000 images includes 250000 labelled images,150 object instances,80 object categories. single shot detector (SSD) achieves better accuracy over dense sampling of object locations and comparable accuracy than RCNN series detectors. Overall this model achieves comparable accuracy[1].G. Chandan et. al proposed a system for real time object detection and tracking using deep learning and OpenCV models such as MobileNet, SSD, Fast-RCNN, Single Shot Detector, Common Objects in Context (COCO), YOLO, CNN, VGG-16. These models can be deployed in CCTVs, drones and other surveillance devices. dataset used here is videos and images. among all these models, Fast-RCNN and SSD gives better accuracy.In case of object detection, it gives 99% accurate results[2].Guansong Pang et. al proposed a system of deep anomaly detection with deviation networks. Models we used here is DevNet which is compared with four methods including REPEN, DSVDD, FSNet, iForest. Dataset used for train and test this model is nine publicly available real-world datasets of images for critical domains like fraud detection, and other four datasets contains real anomalies. Modified DSVDD improves 30% more accuracy achieves by this model which is comparably good accuracy compared to original DSVDD[3].

Ayesha Younis et. al proposed real-time object detection using pre-trained deep learning models such as MobileNet, RCNN, YOLO and SSD. These model combines MobileNet and SSD for fast and efficient deep learning-based method of object detection.14 fps processing speed push from original 6fps. Object images dataset are used here to train the model. The high accuracy object detection procedure has been achieved by using MobileNet and SSD detector for object detection[4].Robert J et. al proposed a system of real time object detection sysetm on mobile devices using various deep learning models like CNN, MobileNetV2, PeleeNet. PeleeNet achieves 76.4% mean average precision on PASCAL VOC2007 dataset and on ImageNet ILSVRC 2012 dataset, these model achieves higher accuracy and also 1.8 times faster speed than MobileNetV2.It contains 14,580 training images and 6000 validation images dataset. PeleeNet achieves 79.25% accuracy on standford dogs dataset and 90.6% accuracy on ImageNet ILSVRC 2012 dataset[5].Clemens-Alexander Brust et. al proposed a system of active learning for deep object detection using various deep learning models such as CNN, YOLO, R-CNN. CNN used for feature extraction, SVM used to score extracted features. Dataset used here is PASCAL VOC 2012 and VOC dataset.It gaining accuracy as fast as possible, by minimize the human supervision. Fast R-CNN offers a improvements in accuracy and speed[6].S A Sumon et. al proposed a system of violence detection by pretrained modules with different deep learning approaches such as three ImageNet models VGG16, VGG19, ResNet50 which is used for feature extraction from the videos frames, CNN, LSTM. In these models 30 frames is extracted from videos at a time. 220 videos dataset used for train and test the model which is collected from social media platforms. ResNet50 gives 97.06% accuracy with LSTM.VGG16, VGG19 & ResNet50 gives 92.34%,93.50% & 97.06% accuracy [7]. M Sharma et. al proposed a system of video surveillance for violence detection using various deep learning models like ResNet50 for feature extraction from video frames, ConvLSTM. Here in these models different violence seen of different situations videos dataset are used like KTH, Hockey fight dataset, violent-flows datasets. Accuracy achieves by ResNet50 is 89.9%, whereas VGG19 not gives considerable accuracy which is 79.3% [8].

K Doshi et. al developed an online anomaly detection technique for surveillance videos utilizing transfer learning, which is utilized for every shot learning used for statistical detection methods' capability and features extraction power of neural network-based models. By using a small number of labeled nominal samples, it provides a technique to discover abnormalities and greatly reduces complexity. The suggested approach can smoothly switch between many-shot and few-shot learning[9]. Arun Akash S A et. al proposed a system to detect human violence using deep learning techniques. The action and attributes of the video are detected and predicted using deep learning algorithms. They did this by using the YOLO-5 model and Inception-v3 models to identify the violent behavior, the number of participants, and the weapons that were causing harm. With a 74% accuracy rate, this model can be used in real time as software or an API[10]. Christian Szegedy et. al proposed a system an object detection model using Deep Neural Network on image classification task. This system presents a straightforward and effective formulation as a regression problem to object bounding box masks. Additionally, the authors define a multi-scale inference procedure. When used with a multi-scale course, it produces good results, but because a network must be trained for each item and mask type, there is a computational cost associated with the training process[11].

F U M ULLAH et. al proposed a system of vision based violence detection in surveillance videos cameras using ANN and machine intelligence, surveillance mechanism and deep learning based method like CNN, RNN(LSTM). Datasets utilized for violence detection are sourced from various platforms, including YouTube, real-time CCTV footage, movies, or recordings captured with mobile phones, and it's achieve different accuracy in different suitution like 89.5% accuracy in violence in movies [12]. N Mohod et. al proposed a system of object detection in surveillance video using YOLOv4 and YOLOv5 techniques. Here, two object detection algorithms—the region-based and the region-free method—are utilized. This model places more emphasis on the region-free approach due to its high detection speed and accuracy. basically, this model is trained using ATM footage from a dataset that was produced and is not publicly available. The accuracy that YOLOv5 achieves is 84%, which is better than the 56% accuracy that YOLOv4 yields [13]. N Mohod et. al proposed a system of human detection in surveillance video using deep learning approach. It is divided into two stages: a one-stage identifier and a two-stage identifier. Two distinct object detection approaches that work well are Mask-RCNN and YOLOv4.In this model, Mask-RCNN outperforms YOLOv4 with an accuracy of 85% whereas YOLOv4 only achieves 65% accuracy [14].

## III. PROPOSED SYSTEM

### A. *Hardware Requirements*

CPU i5 processor, RAM 4GB, OS Windows 11, ROM 250 GB

### B. *Software Requirements*

Python IDE Software, Jupyter, Google Colab

### C. *Methodology*

#### 1) *Datasets*

There are two categories in the dataset: violence and non-violence. Within the root directory, a folder with 5,000 photos represents each category. These photos are different in size, thus before using them to train the models, preprocessing is required. photographs showing street fights and tussles are included in the Violence class, while CCTV point-of-view photographs capturing ordinary life situations are included in the Non-Violence class. The dataset also includes interviews and photos from sporting events to increase variety.

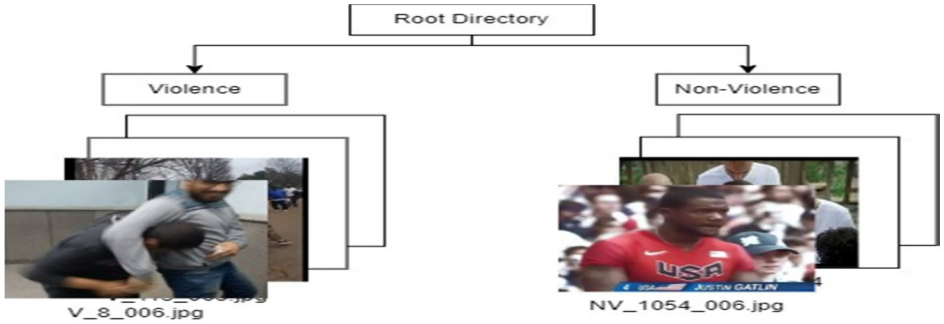



Fig. 1: Dataset structure (comprising of 5000 images for each class: Violence and Non-Violence)

*2) Frame Extraction*

Before moving on to any other activities, preparation is required due to the size variation of the photos. The 'imgaug' Python package is utilized to implement particular preprocessing methods. One of these is a horizontal flip, in which the image is rotated either vertically or horizontally. This helps the model learn invariant features and reduces the possibility that it would overfit to specific object orientations in the images.By multiplying each pixel value within the range of [1, 1.3], random brightness adjustment is applied, which modifies the brightness of the image at random. This modification lessens the model's susceptibility to fluctuations in illumination by strengthening its resistance to changes in lighting conditions. Additionally, each image is subjected to a zooming factor of 1.3 in order to eliminate any unnecessary information.

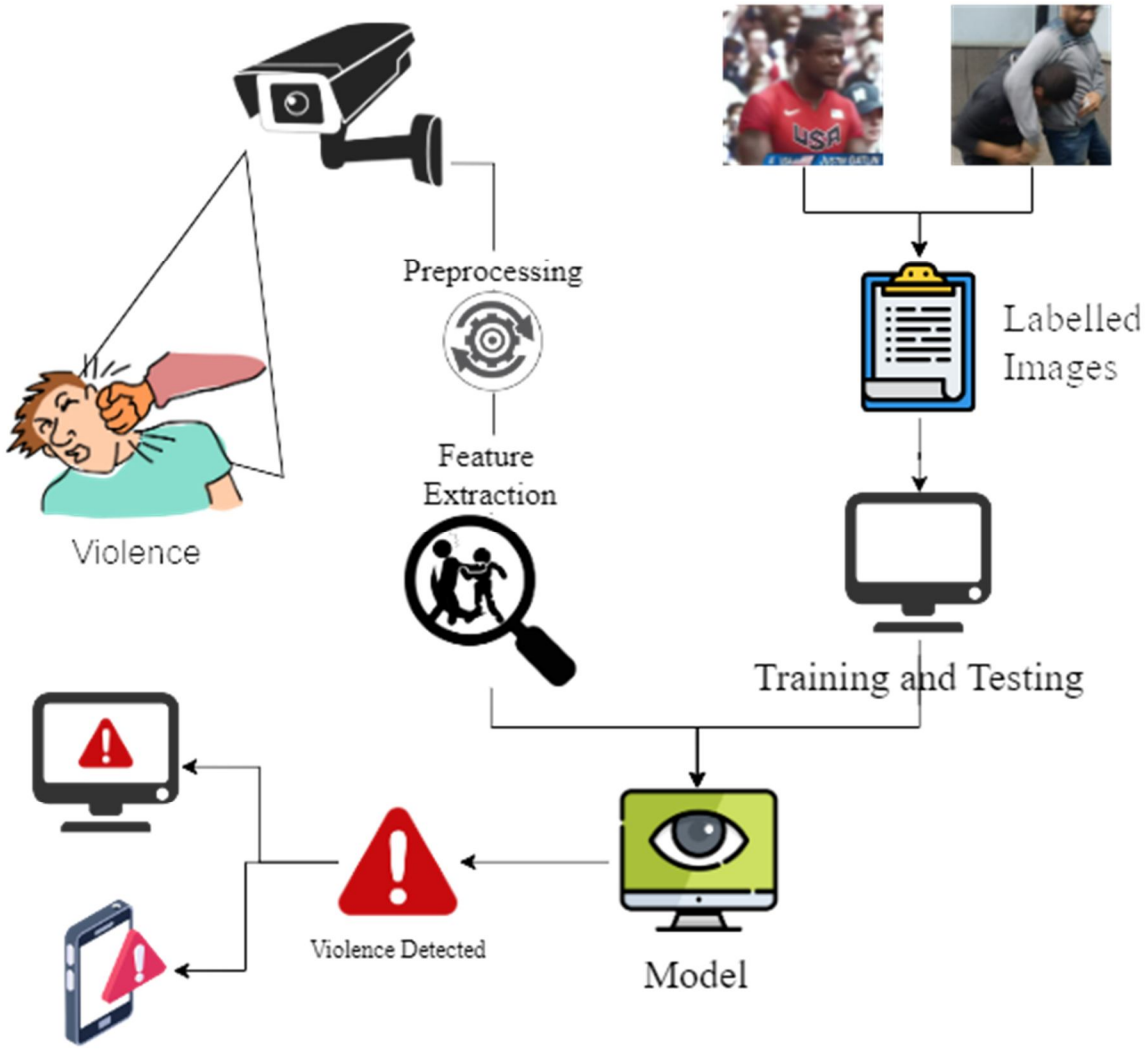



Fig. 2: Architecture of the proposed system.

*3) Feature Extraction*

One of the most important stages in employing deep learning models, such as the ones this study analyzed MobileNetV2, to detect anomalies is feature extraction. It entails converting unprocessed picture input into a more condensed and insightful representation that the model can use to classify images accurately. The aim of this study is to identify characteristics in videos that differentiate between violent and non-violent content. Pre-trained deep learning models, including MobileNetV2, are used in this study to extract features. These models have already acquired strong feature representations for visual content thanks to their extensive training on massive image datasets. By fine-tuning their learnt features on the video anomaly detection task, the models are able to distinguish between violent and non-violent.

*4) MobileNetV2*

A lightweight model that can classify photos in real time is called MobileNetV2. Remaining blocks in Mobilenet-V2 are utilized to upsample and compress image samples/frames into larger groups of images, which are then downsized and converted into 1*1 convolution. The contracting path is located on the left side of MobileNetV2's design, while the classifier head is located on the right. The contracting path consists of a standard convolutional network design with many applications of dual 3*3 convolutions, each followed by 2*2 max pooling operations and ReLU activations to downsample the feature maps. These three processing phases are referred to as a block. Which is carried out multiple times to give the network depth, which leads to a number of completely linked layers at the classifier stage. Every step of the downsampling process doubles the number of feature channels.Computing filters are frequently applied throughout the whole dataset at each layer of the layers that are used to produce the feature maps in order to increase the training efficiency.

The feature map matrix is vectorized at the end of the last block of the MobileNetV2 stage and supplied into the fully connected layer, also known as the classifier stage of the neural network. Activation functions, such as ReLu, are ultimately used to categorize the outputs of violent and non-violent samples.

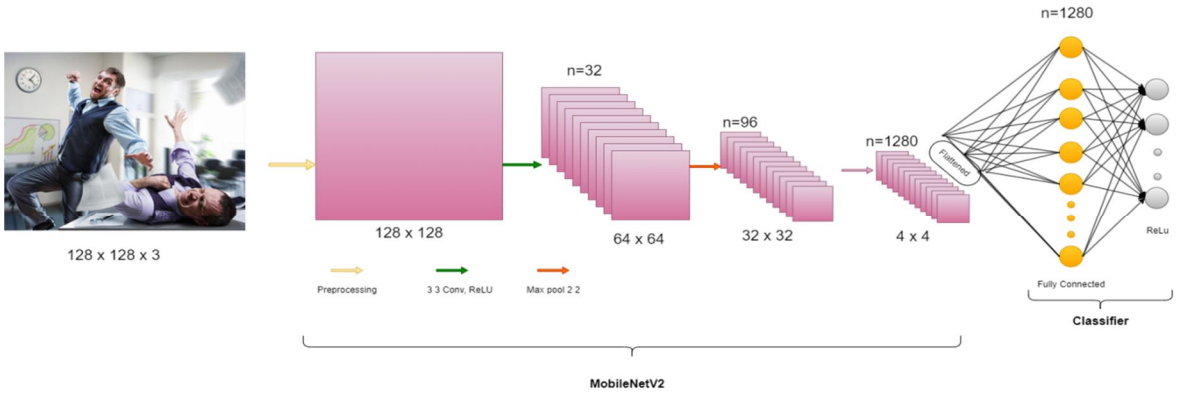



Fig. 3: Architecture of MobileNetV2.

*5) YoloV5s*

The three main components of the YOLOv5 network are the Neck, Head, and Backbone. A convolutional neural network, which bundles and produces picture representational information at differing granularities, makes up the Backbone. The neck of the architecture is made up of several layers that combine and integrate image-representational elements in order to advance prediction. In a similar vein, the head accesses box and class prediction functionality by using features from the neck. The Yolov5 CSPDarknet53 backbone has 29 3 × 3 convolutional layers, a 725 × 725 receptive field size, and a total of 27.6 M parameters. A convolution layer is also included in the down-sampling layer, but it lacks the personality association. That is how it continues for a few deeply layered levels. The final layer, known as normal pooling, creates 1000 element maps for each feature map by normalizing the ImageNet data. The result would be a 1000-dimensional vector, which would then be directly fed into the Softmax layer, making it entirely convolutional. At that point, we would obtain the categorization of the image according to its class.

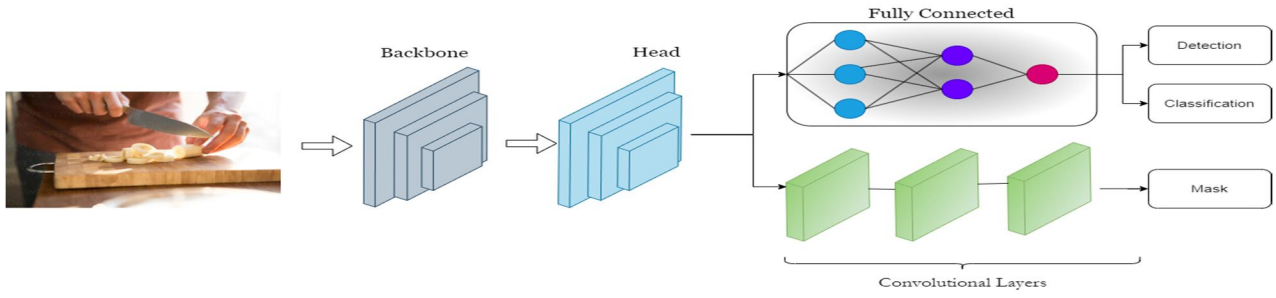



Fig. 4: Architecture of Yolov5s

### *D. Experiments*

Table: Comparision of Train and Test of MobileNetv2 models

| MobileNetv2 | Train | Test |
|---|---|---|
| Accuracy | 93.94 | 89.20 |
| Loss | 17.02 | 25.95 |
| F1-score | 89.00 | 89.00 |

#### *1) Model Training*

Both binary cross-entropy loss and accuracy metrics are used to evaluate each model. Graphs showing accuracy and loss are used to visually portray each model's performance. To evaluate the performance of the models, an accuracy based on the validation portion of the dataset and an overall loss score are calculated. Table 2 presents these findings. There are 75 epochs in all that each model runs. To get the best performance, the hyperparameters must be properly adjusted. The input shape, which is set to 128*128*3, specifies the dimensions of the input images. The step size during parameter updates is controlled by the learning rate, which is 0000.1 for each of the three models. Batch Size, which is set to 64, indicates how many samples are processed per iteration, and epochs, which are set to 75, indicate how many passes the dataset undergoes during training. The optimization process is influenced by Adam, the optimizer that is employed. For each model, the batch normalization technique, ReLU activation, and sigmoid activation function are used.

#### *2) Model Evaluation Metrics*

Our study investigated the effectiveness of MobileNet V2 for anomaly detection in surveillance videos. To assess its ability to distinguish normal activities from anomalies, we employed standard metrics: accuracy, loss, and F1-score. MobileNet V2 achieved an impressive 88% accuracy on the test set, signifying a strong capability to differentiate between usual occurrences and anomalies in the footage. This translates to the model effectively identifying deviations from established patterns, pinpointing anomalies with a high degree of success. Additionally, the model achieved a low loss of 26, indicating a close match between its predictions and the actual labels. This demonstrates the model's efficient learning of patterns, enabling it to effectively distinguish between normal and anomalous behaviors. Furthermore, the F1-score of 88 on the test set highlights a crucial aspect for anomaly detection - a balanced performance. The F1-score considers both precision (correctly identified anomalies) and recall (identifying all actual anomalies). A high F1-score signifies the model avoids over-predicting anomalies (high precision, low recall) and under-predicting them (low precision, high recall). This balance is essential to minimize false alarms while ensuring true anomalies are captured. MobileNet V2's F1-score underscores its ability to achieve this crucial balance. The combined evaluation using accuracy, loss, and F1-score paints a promising picture for MobileNet V2 in anomaly detection for surveillance videos. The high accuracy indicates strong classification ability, the low loss signifies effective learning, and the balanced F1-score highlights the model's ability to identify anomalies accurately. These results suggest that MobileNet V2 is a viable option for anomaly detection tasks in surveillance applications. While MobileNet V2 demonstrates promising results, further exploration holds value. Investigating techniques to improve upon the current performance, such as hyperparameter tuning or incorporating additional training data, could lead to even better anomaly detection capabilities.

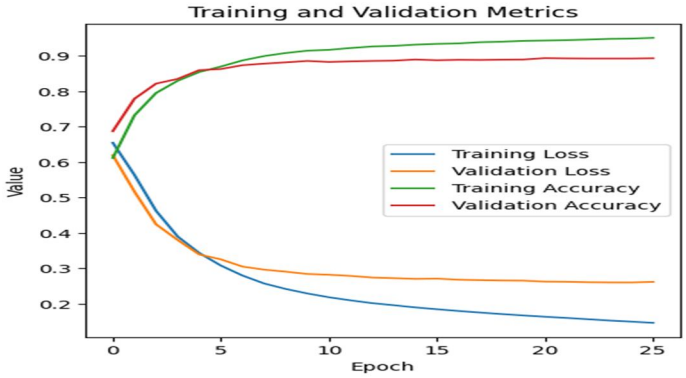



Fig. 5: MobileNetv2 Training and Validation Accuracy and Loss

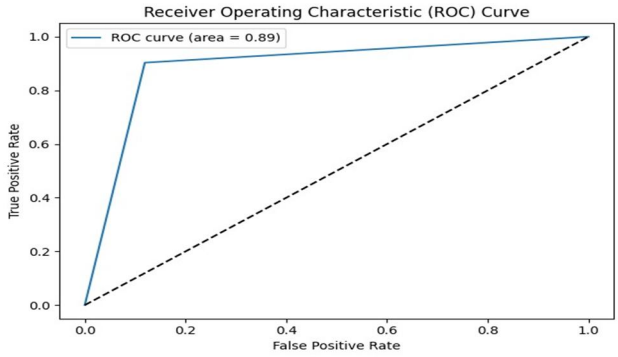



Fig. 6: MobileNet ROC and AUC Score

The ROC curve in the image visualizes MobileNet V2's performance in anomaly detection. The curve's proximity to the upper left corner indicates good ability to separate normal and anomalous video frames. While specific points on the curve represent TPR (correctly identified anomalies) versus FPR (normal events misclassified), the high AUC of 0.89 signifies the model's overall strength in differentiating between them. This AUC suggests MobileNet V2 can accurately rank true anomalies higher than normal events, making it a promising choice for anomaly detection in surveillance applications.

## IV. OUTPUT IMAGES

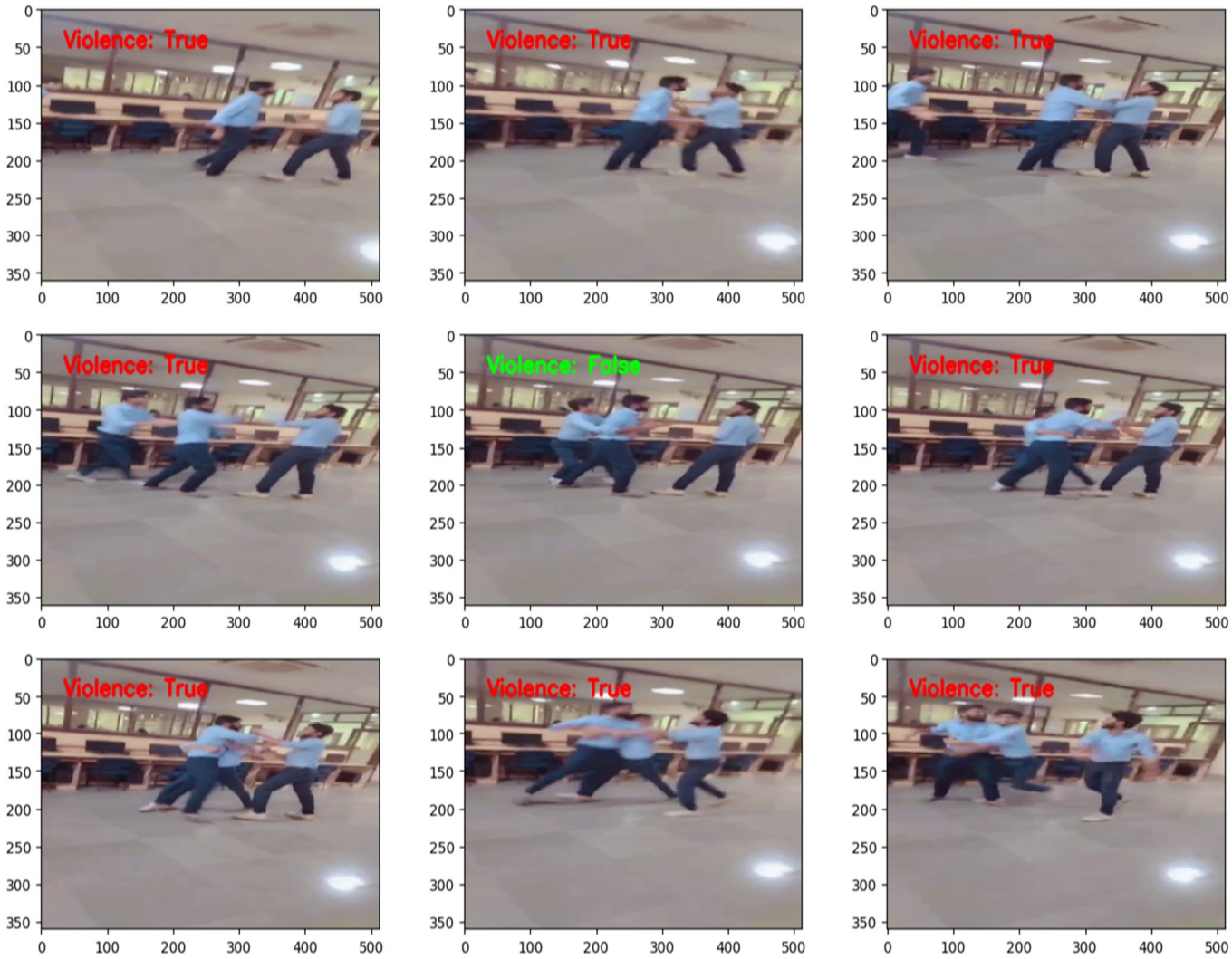



Fig. 6: Output: MobileNetV2

## V. CONCLUSION

In conclusion, Deep learning-based anomaly detection offers a significant advancement in video surveillance. Our proposed system, leveraging the MobileNetV2 architecture, has achieved strong and consistent performance in detecting anomalies within surveillance footage. Through comprehensive testing, the model attained an accuracy of 88% with a loss of 26, demonstrating its effectiveness. Additionally, the F1-Score of 89% indicates a low rate of false alarms, ensuring the system's reliability. This approach prioritizes anomaly detection over specific anomaly classification or precise timestamp identification. By integrating this high-performance model into surveillance systems, we anticipate a substantial reduction in human error and manual workload, leading to a more efficient and effective security monitoring system.

## REFERENCES

[1] Murthy, C.B.; Hashmi, M.F.; Bokde, N.D.; Geem, Z.W. Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review. *Appl. Sci.* **2020**, *10*, 3280. https://doi.org/10.3390/app10093280

[2] G. Chandan, A. Jain, H. Jain and Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2018, pp. 1305-1308, doi: 10.1109/ICIRCA.2018.8597266

[3] Guansong Pang, Chunhua Shen, and Anton van den Hengel. 2019. Deep Anomaly Detection with Deviation Networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). Association for Computing Machinery, New York, NY, USA, 353–362. https://doi.org/10.1145/3292500.3330871

[4] Ayesha Younis, Li Shixin, Shelembi Jn, and Zhang Hai. 2020. Real-Time Object Detection Using Pre-Trained Deep Learning Models MobileNet-SSD. In Proceedings of 2020 6th International Conference on Computing and Data Engineering (ICCDE '20). Association for Computing Machinery, New York, NY, USA, 44–48. https://doi.org/10.1145/3379247.3379264

[5] Wang, R.J., Li, X. and Ling, C.X., 2018. Pelee: A real-time object detection system on mobile devices. Advances in neural information processing systems, 31.

[6] Brust, C., Käding, C., & Denzler, J. (2018). Active Learning for Deep Object Detection. ArXiv, abs/1809.09875.

[7] Sumon, S.A., Goni, R., Hashem, N.B., Shahria, T. and Rahman, R.M., 2020. Violence detection by pretrained modules with different deep learning approaches. Vietnam Journal of Computer Science, 7(01), pp.19-40.

[8] Sharma, M., Baghel, R. (2020). Video Surveillance for Violence Detection Using Deep Learning. In: Borah, S., Emilia Balas, V., Polkowski, Z. (eds) Advances in Data Science and Management. Lecture Notes on Data Engineering and Communications Technologies, vol 37. Springer, Singapore. https://doi.org/10.1007/978-981-15-0978-0_40

[9] Doshi, K. and Yilmaz, Y., 2020. Any-shot sequential anomaly detection in surveillance videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 934-935).

[10] Akash, S.A., Moorthy, R.S.S., Esha, K. and Nathiya, N., 2022, August. Human violence detection using deep learning techniques. In Journal of Physics: Conference Series (Vol. 2318, No. 1, p. 012003). IOP Publishing.

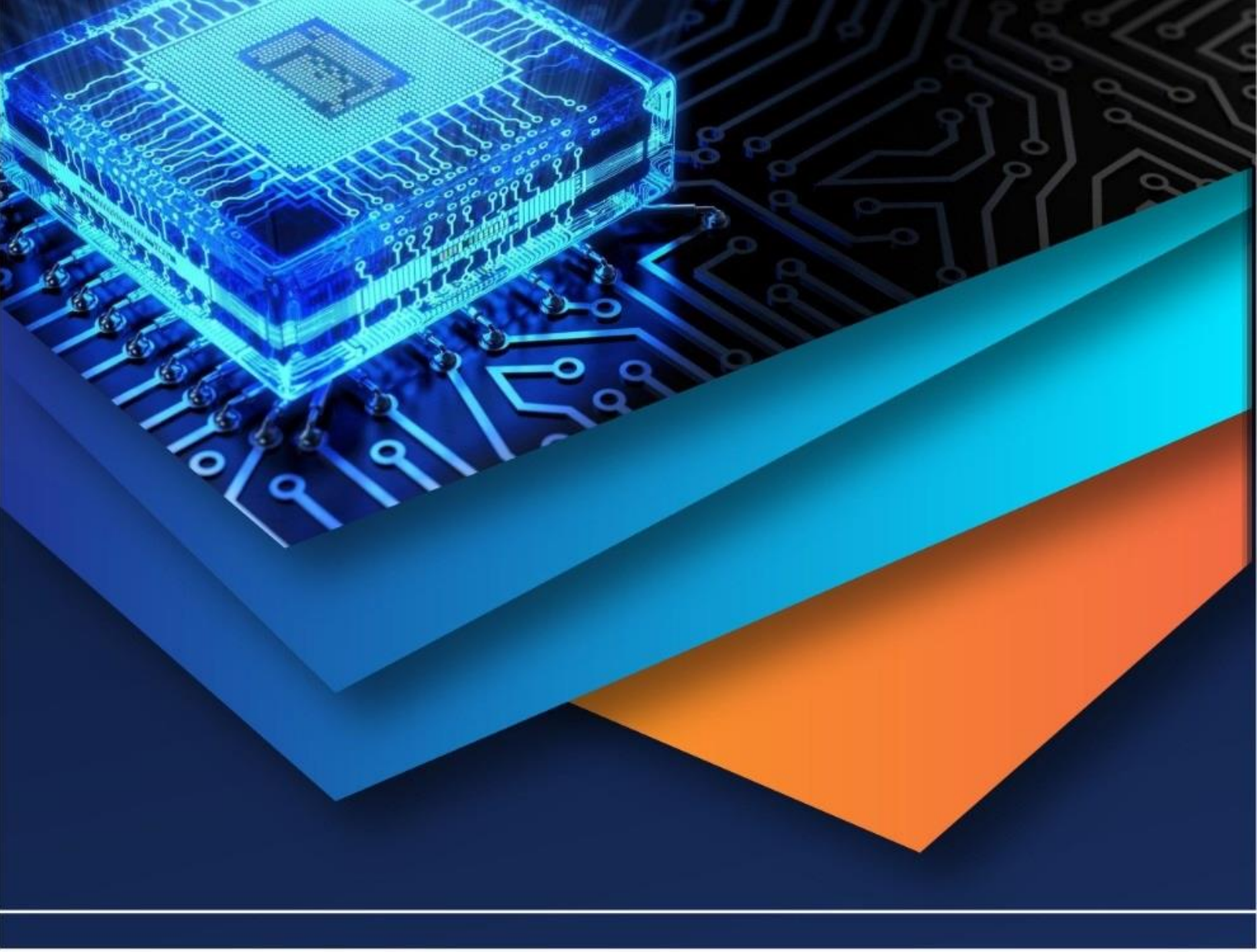
[11] Szegedy, C., Toshev, A. and Erhan, D., 2013. Deep neural networks for object detection. Advances in neural information processing systems, 26.

[12] Ullah, F.U.M., Obaidat, M.S., Ullah, A., Muhammad, K., Hijji, M. and Baik, S.W., 2023. A comprehensive review on vision-based violence detection in surveillance videos. ACM Computing Surveys, 55(10), pp.1-44.

[13] Mohod, N., Agrawal, P., Madaan, V. (2023). YOLOv4 Vs YOLOv5: Object Detection on Surveillance Videos. In: Woungang, I., Dhurandher, S.K., Pattanaik, K.K., Verma, A., Verma, P. (eds) Advanced Network Technologies and Intelligent Computing. ANTIC 2022. Communications in Computer and Information Science, vol 1798. Springer, Cham. https://doi.org/10.1007/978-3-031-28183-9_46.

[14] N. Mohod, P. Agrawal and V. Madan, "Human Detection in Surveillance Video using Deep Learning Approach," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-6, doi: 10.1109/ISCON57294.2023.10111951.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY