International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 08 | Aug 2024 www.irjet.net p-ISSN: 2395-0072

Human Suspicious Activity Detection using Machine Learning

Joshitha Lakshminarayana[1], Bhoomika K[2], Keerthana M[3], Dr.Vrunda Kusanur[4]

[1]Dept. of ECE BNM Institute of Technology Bengaluru, India,
[2]Dept. of ECE BNM Institute of Technology Bengaluru, India,
[3]Dept. of ECE BNM Institute of Technology Bengaluru, India,
[4]Associate Professor, Dept. of Electronics and Communication Engineering, BNM Institute of Technology, Karnataka India

---

Abstract--*Current urban environments require advanced surveillance systems that can identify unusual activity in densely populated regions in order to protect public safety and security. This project uses an advanced multi-model approach to detect anomalous behavior in pedestrians, in order to meet this need. The combination of ResNet50, Inception V3, and the Slow Fast model represents an all-encompassing approach that utilizes deep learning to identify minor variations achieving an accuracy of 94%. By improving the effectiveness of surveillance systems in public areas, this project hopes to contribute to the general safety and security by offering an effective way of identifying and handling anomalous behavior.*

Key words—Anomalous behavior, Pedestrians, ResNet50, Inception V3, Slow Fast model, Deep learning, Advanced multimodel approach, Accuracy (94%), Public safety.

1.INTRODUCTION

In today's world, abnormal activity indicates threats and risks to others. An anomaly can be defined as something that deviates from what is expected, common, or normal. Because it is difficult to continuously monitor public spaces and classify them, intelligent video surveillance is necessary. So choosing an appropriate framework for identifying suspicious activity plays a very significant role.

Abnormal activity detection, monitoring and identifying abnormal conditions is a very difficult task. In today's life, the use of security cameras is increasing due to serious crimes. For some the performance is incredible, for others it is not. Therefore, choosing an appropriate framework to identify suspicious activities plays an important role . New monitoring systems use deep learning models to identify malicious programs. Deep learning is very popular and is a type of machine learning.

It is popular because it can process unnecessary information. Deep learning utilizes and utilizes deep learning neural networks. He has good calculations. It also provides great flexibility when it needs to handle many features. These features are taken from data and do not cause any problems. The deep learning model has many layers through which it passes data. Each layer comes with features derived from non-standard data. This feature is transferred to the next layer of the network.

Low level features are extracted from the first layer and the following layers will provide these features. Creates a complete feature representation. It solves many problems that arise in traditional analysis. Provides better performance by deep neural networks.

Research literature has gathered the following point's supports the use of deep learning in analysis Analyzing human experience activity by observing video is a frequently conducted research in image processing and computer vision operates in its field.

Bus stations, train stations, airports, banks, shops, homes, schools and colleges, car parks, roads, etc. through visual monitoring. Human activities in public areas such as can be monitored. Prevent violence, theft, accidents and illegal parking, torture, fighting, theft, crime and other activities no. It is very difficult to constantly monitor public places, so there is a need for intelligent video surveillance that can monitor human activities and distinguish them into normal and abnormal; and can generate an alarm.

2.SYSTEM ARCHITECTURE

2.1 Inception V3:

Introduction: Inception V3 stands as a pinnacle achievement in the realm of convolutional neural networks (CNNs), a product of continuous refinement within the Inception series pioneered by Google. This state-of-the-art architecture is specifically crafted for image classification and object recognition, embodying a synthesis of innovative design principles that elevate its performance in comparison to its predecessors.

Architecture Overview: The evolution of Inception V3 is marked by a departure from conventional approaches. The architecture comprises a meticulously designed ensemble of convolutional modules, each tailored for a specific purpose. These modules, including 1x1, 3x3, and 5x5 convolutions, collectively enable the network to capture features at varying scales. Furthermore, the introduction of parallel pathways, often referred to as Inception modules, allows simultaneous feature extraction at different receptive field sizes. This parallelism empowers the model to discern fine details alongside broader contextual information, contributing to its robust recognition capabilities.

Efficient Feature Extraction: At the core of Inception V3's success lies its ability to perform efficient feature extraction. By diversifying convolutional filters, the model excels in capturing intricate details crucial for image understanding. The spatial hierarchies created through the amalgamation of different filters contribute to the network's capacity to comprehend complex structures and patterns. Notably, the incorporation of 1x1 convolutions strategically reduces the dimensionality of feature maps, striking a balance between computational efficiency and information retention.

Methodology -: Inception V3 goes beyond conventional architectures by introducing auxiliary classifiers during training. These classifiers, located at intermediate layers, serve a dual purpose. While aiding in the training process by providing additional paths for gradient flow, they also contribute to the model's overall robustness. This innovative approach mitigates issues such as the vanishing gradient problem, promoting better convergence during the training phase.

Versatility and Transfer Learning: The versatility of Inception V3 extends to its applicability in transfer learning. Pre-trained models can be leveraged across diverse domains, offering a significant advantage in scenarios with limited labeled data. This adaptability has rendered Inception V3 a popular choice for a wide array of computer vision tasks, contributing to breakthroughs in image recognition and classification.

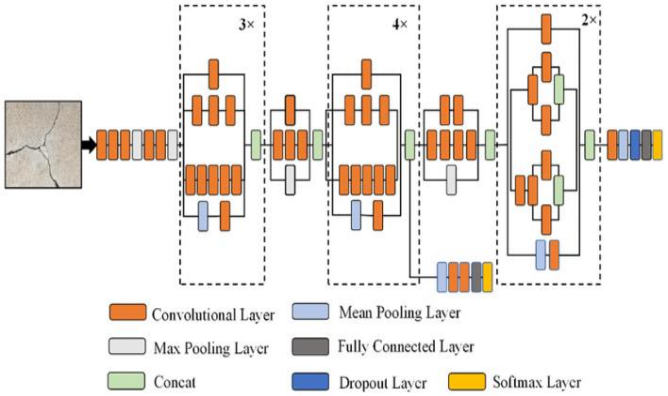

Fig. 1: The architecture of Inception-V3 mode

2.2 Resnet-50:

Introduction: ResNet-50, a convolutional neural network (CNN) developed by Microsoft researchers, has gained acclaim for its deep learning capabilities in image recognition. It effectively tackles challenges related to training deep networks through the introduction of residual connections.

Residual Learning and Skip Connections: ResNet's key innovation lies in residual blocks featuring skip connections. Traditional deep networks grapple with the vanishing gradient problem, hindering training of extensively layered models. Residual blocks mitigate this by learning the residual mapping, the input output difference, and combining it with the original input through a skip connection, easing optimization.

Architecture Overview: Comprising 50 layers, ResNet-50 includes convolutional layers, batch normalization, activation functions, and fully connected layers. The architecture is segmented into stages with multiple residual blocks, incorporating global average pooling and a fully connected layer for classification.

Convolutional Layers: The initial layers serve as feature extractors, capturing low-level features like edges. Progressing through stages, the convolutional layers learn intricate features, enabling discrimination between diverse patterns.

Residual Blocks: Central to ResNet's success, each block contains convolutional layers, batch normalization, and ReLU activation. The skip connection facilitates learning residual features, capturing fine-grained details. Global Average Pooling: Replacing fully connected layers, global average pooling reduces feature dimensionality, enhancing interpretability and generalization to new data. Pre-Trained Weights and Transfer Learning: ResNet-50 is often pre-trained on datasets like ImageNet, leveraging transfer learning for specific applications. This approach saves computational resources and time.

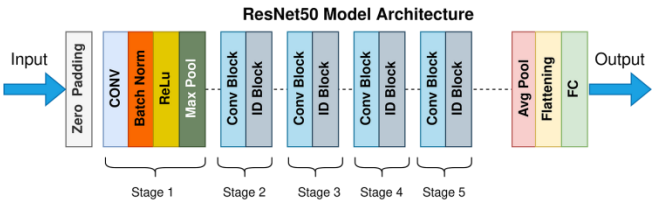

Fig. 2: ResNet50 Model Architecture

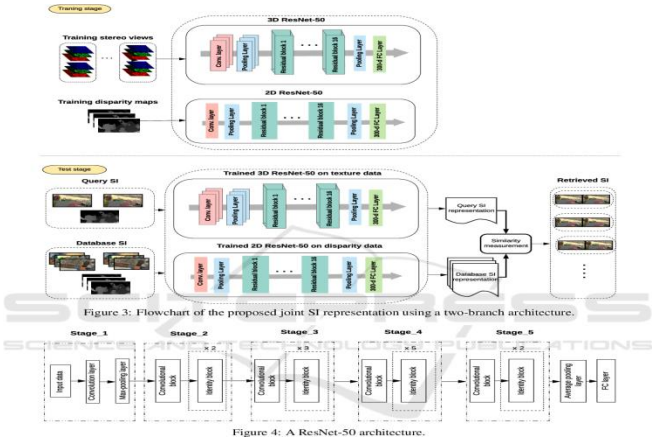

Fig. 3: ResNet50 Architecture

2.3 Slow-Fast Model

Introduction: The SlowFast model represents a breakthrough in the domain of video understanding, addressing the inherent challenges posed by the temporal dimension in visual data. Introduced as a novel architecture by Facebook AI Research (FAIR), the SlowFast model is designed to capture both high and low-frequency temporal information efficiently. This innovative approach is particularly impactful in video recognition tasks, where a nuanced understanding of temporal dynamics is essential for accurate analysis.

Architecture Overview: At the heart of the SlowFast model is a dual-path architecture, featuring two separate streams for processing video frames at different temporal resolutions. The "Slow" pathway operates at a reduced frame rate, capturing the broader context and global temporal information. Conversely, the "Fast" pathway processes frames at a higher frame rate, focusing on finer details and rapid temporal changes. This dual-stream design allows the model to balance the trade-off between capturing long-term dependencies and reacting swiftly to temporal nuances, resulting in a more comprehensive understanding of video content.

Efficient Feature Extraction: The efficiency of the SlowFast model is underpinned by its adept feature extraction process. The Slow pathway, with its reduced frame rate, produces high-level semantic features, capturing the overarching temporal structure. Simultaneously, the Fast pathway excels in capturing fine-grained details and subtle temporal variations. The fusion of these pathways, achieved through a carefully designed network architecture, ensures that the model synthesizes a holistic representation that encapsulates both global and local temporal features.

Methodology - Dual-Path Temporal Convolution: A distinctive feature of the SlowFast model is the integration of a dual-path temporal convolution mechanism. This innovation involves separate 3D convolutions for each pathway, allowing the model to process temporal information independently in both the Slow and Fast streams. The combination of these convolutions ensures that the model can effectively capture temporal dependencies across multiple time scales. This methodology enhances the model's ability to discern complex temporal patterns, making it well-suited for a wide range of video analysis tasks.

Spatial-Temporal Fusion: A key strength of the SlowFast model lies in its emphasis on spatial-temporal fusion. The model seamlessly integrates information from both the Slow and Fast pathways through carefully designed fusion layers. This fusion process enables the model to create a unified representation that leverages the strengths of both streams, resulting in a more nuanced and comprehensive understanding of the temporal dynamics within the video data.

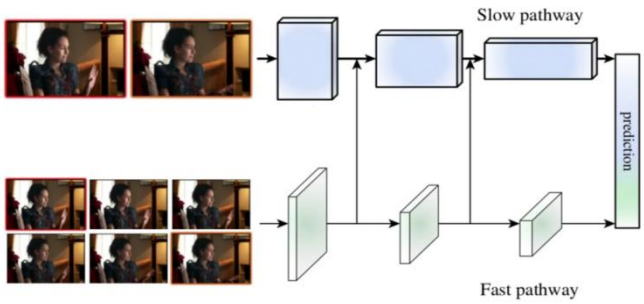

Fig. 4: Illustration of the SlowFast Network with Parameters

3.PROPOSED METHOD

3.1 Introduction

The objective of this project is to create a supervised ML model for detecting abnormal activities , Classify them and notify the respected higher authorities. Fig. 5 represents the complete methodology of the model.

Machine Learning Model for Human Suspicious Activity:

1) Data Collection and Preprocessing: The machine learning model is based on a carefully selected dataset. We started the process of gathering data, collecting a wide range of short videos of abnormal activities detected in public places , from several websites and apps such as kaggle , which included datasets of several classes of abnormal activities such as shoplifting , Stealing, Fighting , Explosion and Vandalism.

Before preprocessing, First the video must be converted into several video frames. In order to ensure a model with optimal performance, a variety of preprocessing processes have been performed to the video frames obtained. These operations consist of resizing each frames , normalizing pixel values, and standardizing image sizes. While normalization improves model convergence, standardization guarantees consistency.

2) Selection and Training of Machine Learning Model: In the quest for architecture conducive to video classification tasks, we opted to train three Machine Learning Models: Slow-Fast, Inception-v3 and Resnet50.

3) Web Application Integration: The trained ML Model is integrated into the web application. A simple interface is created using HTML and CSS to facilitate the usage of ML Model for Detection of abnormal activities in public places. The model is configured to receive video inputs through the website, enabling users to obtain binary classification outputs by uploading videos onto the platform.

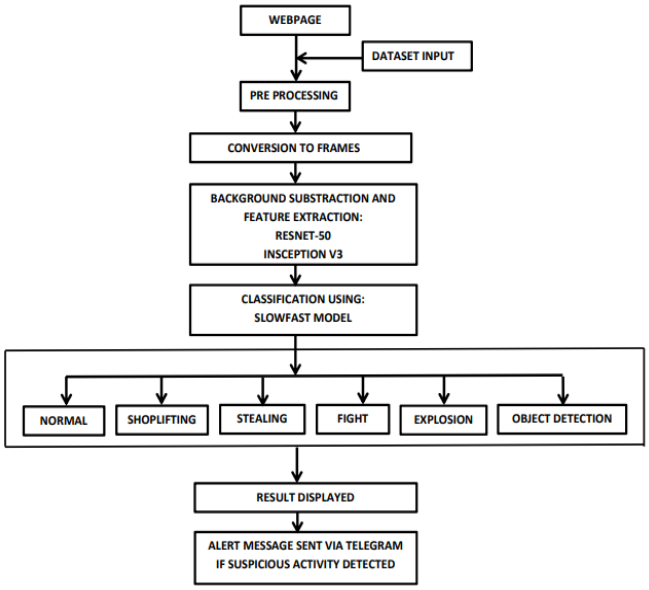

Fig. 5: Flowchart depicting the model

4.RESULTS

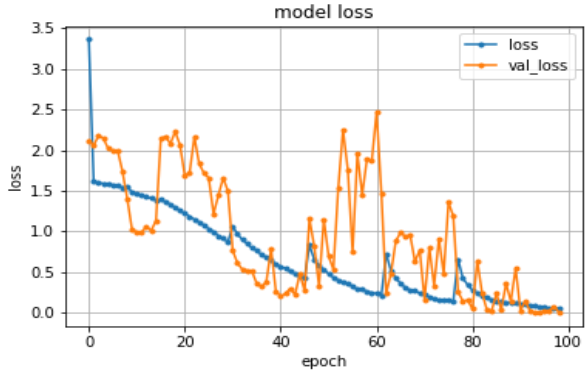

Chart -1: Graph depicts Model Loss

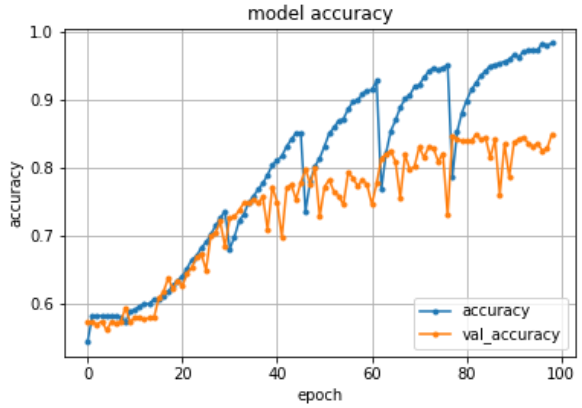

Chart-2: Graph depicts Model accuracy

We plot the model's training and validation accuracy and loss, as shown in Chart -1 and Chart -2. The epochs of the model accuracy and loss for video classification are around 98%. The classification report of the model using the unseen test data is presented in Fig 6. The support values show the number of videos related to each class in the test data.

Table 1: Accuracy and loss values for first 10 epochs

Epoch	Accuracy	Loss	Val_accuracy	Val_loss
1	0.54480773	3.36456636	0.572936654	2.120951414
2	0.58258516	1.61855216	0.572936654	2.057857275
3	0.58300865	1.60326251	0.570537448	2.178224564
4	0.58266985	1.59072744	0.574856043	2.150344133
5	0.58334744	1.58567077	0.561900198	2.024450302
6	0.5822463	1.57488155	0.573416531	1.996793985
7	0.5821616	1.5607918	0.572456837	1.999437571
8	0.5809758	1.54315505	0.574376225	1.742306828
9	0.57403016	1.55894271	0.594529748	1.405586243
10	0.5891919	1.48105642	0.573416531	1.024970889

Table 2: accuracy and loss values for last 10 epochs

Epoch	Accuracy	Loss	Val_accuracy	Val_loss
90	0.9595968	0.11509219	0.787428021	0.547752619
91	0.9666271	0.09904027	0.83733207	0.027833272
92	0.9629849	0.10946636	0.842610359	0.143712908
93	0.9719634	0.08682171	0.845489442	0.020596577
94	0.97272575	0.08332348	0.836852193	0.004468319
95	0.97467387	0.07134484	0.832053721	0.001753665
96	0.9737422	0.07556306	0.835892498	0.023957286
97	0.9837371	0.05333064	0.825815737	0.019642983
98	0.9806031	0.05738091	0.82869482	0.078213073
99	0.9839065	0.05292019	0.850287914	0.006304689
100	0.95400643	0.13028792	0.761036456	0.356913626

Fig. 6: Figure depicts normal case

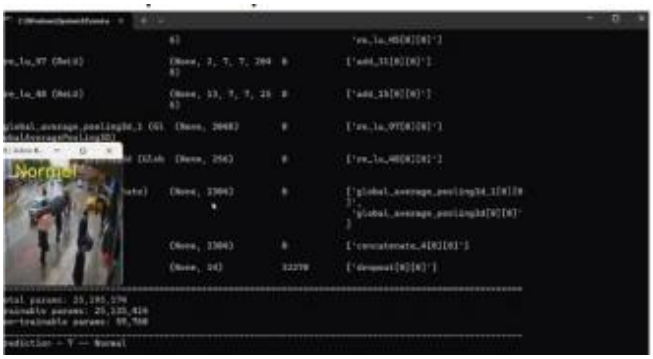

Fig. 7: Figure depicts Normal Prediction

Fig. 8: Figure depicts Shoplifting Case

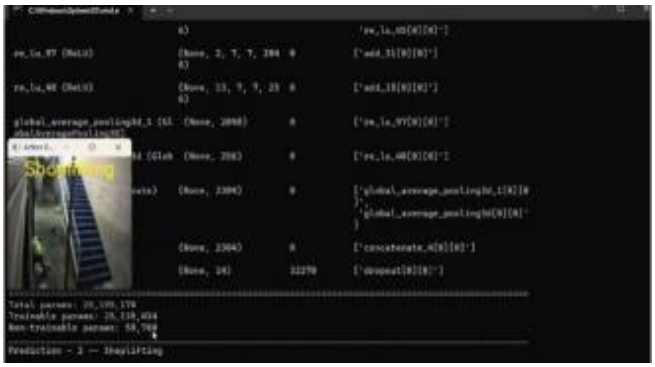

Fig. 9: Figure depicts Shoplifting Prediction

Fig. 10: Figure depicts Fighting Case

Fig. 11: Figure depicts FIghting Prediction

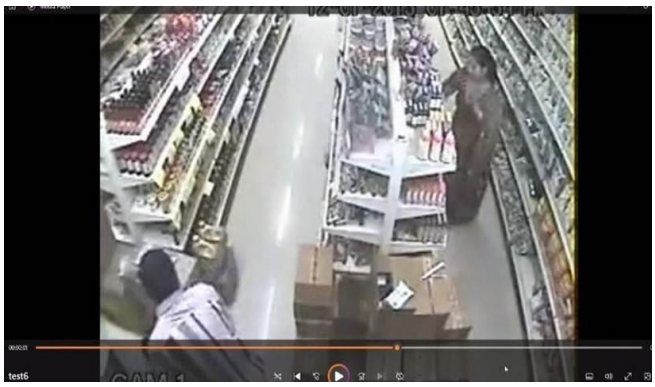

Fig. 12: Figure depicts Stealing/Robbery Case

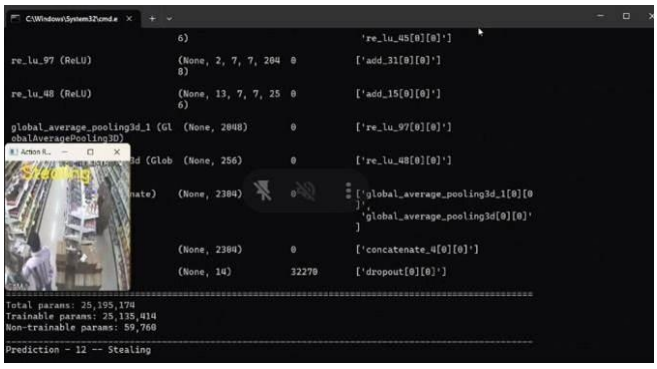

Fig. 13: Figure depicts Stealing/Robbery Prediction

Fig. 14: Alert Message via Telegram

4. CONCLUSION AND FUTURE SCOPE

Machine learning-based detection of suspect human activity is a major step forward for improving security protocols in a variety of sectors. We may examine enormous volumes of data to find patterns and abnormalities that can point to potentially dangerous or illegal activity by utilizing the capabilities of machine learning algorithms.

In this regard, machine learning's capacity for adaptation and continuous improvement is one of its main advantages. By means of ongoing exposure to labeled data, these algorithms are able to enhance their comprehension of typical behavior in contrast to potentially suspicious activities. The creation of more reliable and precise detecting systems is made possible by this versatility.

However, the caliber and variety of the training data greatly influences how successful these machine learning models are. Potentially producing false positives or negatives, biases or shortages in the training data might result in distorted or incomplete representations of suspicious behavior. Consequently, in order to guarantee the generalizability and reliability of the model, it is crucial to collect datasets that cover abroad variety of events and demographics. Furthermore, there are major moral issues raised by the use of machine learning for suspicious behavior identification, especially with regards to privacy and individual rights. It is important to pay close attention to matters like data anonymization, authorization, and transparency in algorithmic decision-making in order to achieve a balance between the necessity for security and respect for individual liberties.

REFERENCES

[l] Abnormal Behavior Detection in Video Surveillance Using Inception-v3 Transfer Learning Approaches. Sabah AbdulazeezJeburl , Khalid A. Hussein2 , HaiderKadhim Hoomod3 JUNE 2023.

[2] .Human Suspicious Activity Detection using Deep Learning RachanaGugale, AbhiruchiShendkar, ArishaChamadia , Swati Patra, DeepaliAhir. June 2020.

[3]Pedestrian Detection and Tracking using HOG and Kalman Filter" byS. S. Patil and S. S. Pawar (2016).

[4]Z. Kain, A. Y. Ouness, I. E. Sayad, S. Abdul-Nabi and H. Kassem, "Detecting Abnormal Events in University Areas," 2018 International Conference on Computer and Applications (ICCA), Beirut, Lebanon , 2018, pp. 260-264, doi: 10.1109/COMAPP.2018 .8460336.

[5]A. B. A., P. P. and V. S., "Detection of Suspicious Human Activity based on CNN-DBNN Algorithm for Video Surveillance Applications ," 2019 Innovations in Power and Advanced Computing Technologies(i-PACT), Vellore, India, 2019, pp. 1-7, doi: 10.1109/i-PACT44901 .2019.8960085.

[6]Anomaly Detection in Video Surveillance using SlowFast Resnet-50 Mahasweta Joshi l Computer Engineering Department CSPIT, CHARUSAT University, Changa, India Jitendra Chaudhari2 Electronics and Communication Department CSPIT, CHARUSAT University,Changa, India.

[7]Pedestrian Detection Using Deep Learning: A Comprehensive Survey" by V. Singh andR.S. Khatkar.

[8] Pedestrian Movement Detection using Video Surveillance: A Comprehensive Review" byE. K. Abagun , et al. (2019).