iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70460

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

CrimeSense: A High-Accuracy Video Crime Classification System

Shekapuram Lokesh Chandra[1], Jagarlamudi Adithya[2], Vecha Sai Badarinadh[3]
School of Computer Science and Engineering, VIT-AP University

Abstract: In this paper, a deep learning based automated crime classification system called CrimeSense is proposed. CrimeSense employs a transfer learning approach utilizing a MobileNetV2 base model fine-tuned with additional convolutional layers to extract robust features from video frames. This approach facilitates the classification of various criminal activities depicted in videos. The system achieves a remarkable accuracy of 97.8% on the UCF Crime Dataset, demonstrating its potential as a useful instrument for law enforcement and other stakeholders in video-based crime analysis. In this paper, a deep learning based automated crime classification system called CrimeSense is proposed.
Keywords: Crime classification, deep learning, video analysis, convolutional neural networks (CNNs), MobileNetV2, transfer learning, UCF Crime Dataset

I. INTRODUCTION

The abundance of video information in today's digital age has created a critical need for effective techniques for identifying and categorizing instances of criminal activity. Because manual video analysis is laborious and prone to mistakes, automated methods are required for the prompt detection and prevention of crimes.

In today's rapidly evolving digital landscape, the ubiquitous presence of surveillance cameras and the exponential growth of video content have transformed the landscape of law enforcement and public safety. Video footage serves as a crucial source of information for investigating crimes, identifying perpetrators, and ensuring the security of communities. However, the sheer volume of video data generated daily poses significant challenges for law enforcement agencies and security personnel in effectively analyzing and classifying instances of criminal activities.

Manual video analysis is labor-intensive and prone to human mistake, which makes it unsuitable for managing extensive surveillance systems. Furthermore, complicated analytical methods that can identify minute patterns and behaviors suggestive of criminal intent are required because to the diversity and complexity of criminal acts documented in video footage, which range from violent attacks to minor thefts.

Existing methods of automated video crime classification have made significant strides in leveraging machine learning techniques to assist in the identification and categorization of criminal activities. But conventional methods frequently depend on manually designed elements and oversimplified algorithms, which could find it difficult to capture the subtleties of criminal conduct found in actual situations. Moreover, these methods may lack the scalability and adaptability required to cope with the ever-changing nature of criminal activities and evolving surveillance technologies.

In light of this, there is an urgent need for sophisticated systems that can independently analyse enormous volumes of video data with great precision and effectiveness. Such systems can not only aid law enforcement agencies in detecting and preventing crimes but also facilitate proactive measures to enhance public safety and security.

CrimeSense, the system proposed in this paper, is a major advancement in tackling the problems associated with automated video crime classification. Through the utilization of deep learning methodologies, CrimeSense aims to provide law enforcement agencies and security personnel with a robust and reliable tool for identifying and classifying criminal activities depicted in video footage. By leveraging state-of-the-art convolutional neural networks (CNNs) and transfer learning approaches [1], CrimeSense endeavors to overcome the limitations of traditional methods and achieve unprecedented levels of accuracy and generalizability in crime detection.

In summary, the proliferation of video content in modern surveillance systems necessitates innovative solutions for automating the analysis and classification of criminal activities. CrimeSense, with its advanced deep learning architecture and tailored training methodology, offers a promising avenue for enhancing the effectiveness of video-based crime analysis and bolstering the efforts of law enforcement agencies in community security and safety.

II. LITERATURE SURVEY

An easy way to comply with IJRASET paper formatting requirements is to use this document as a template and simply type your text into it.

Automated crime classification in videos has emerged as a pivotal area of research with profound implications for law enforcement and public safety. This section delves into the exploration of various models and methodologies undertaken in the development of CrimeSense, shedding light on the iterative process of model selection, training, and refinement to achieve the desired level of accuracy in crime detection.

A. Traditional Machine Learning Methods:

Early endeavors in automated video crime classification primarily relied on traditional machine learning techniques, such as Support Vector Machines (SVMs) [2], coupled with handcrafted features extracted from video frames. While these methods exhibited moderate success, achieving accuracies ranging from 40% to 55.5%, They frequently failed to convey the complex dynamics and patterns present in actual crime scenes. The shortcomings of these methods highlighted the need for more advanced and flexible techniques to address the difficulties associated with video surveillance crime detection.

B. Transition to Deep Learning

Convolutional Neural Networks (CNNs) have transformed computer vision by providing powerful tools for analyzing and interpreting image data [3][4]. There has been some study linking anomaly detection to high dimensionality issues, either directly or indirectly to specifically address specific imaging issues, a number of recent techniques recommend CNN [5][6]. After it was realized that deep learning could revolutionize activities related to video processing, the focus of research turned to using convolutional neural networks (CNNs) [7] for the purpose of classifying crimes. Initially, experiments were conducted with DenseNet121, a widely-used CNN architecture renowned for its depth and parameter efficiency. While DenseNet121 yielded a respectable accuracy of 55.5%, further refinement was deemed necessary to enhance the system's performance.

Subsequently, InceptionV3 [8], another state-of-the-art CNN model renowned for its architecture optimized for image classification tasks, was employed in the quest for improved accuracy. However, the transition to InceptionV3 led to a decrease in accuracy, with the model achieving only 40% accuracy on the dataset. This setback underscored the intricate interplay between model architecture, training methodology, and dataset characteristics in determining the efficacy of crime detection systems.

C. Advancement with MobieNetV2

Undeterred by the initial setbacks, the exploration continued with MobileNetV2 [9] [10], a lightweight CNN architecture designed for mobile and embedded vision applications. Leveraging MobileNetV2 as the foundation, initial experiments demonstrated promising results, with the model achieving a commendable accuracy of 71%. This marked a significant improvement over previous attempts and underscored the importance of selecting appropriate model architectures tailored to the requirements of the task at hand.

D. Fine-tuning and Customization

Recognizing the potential for further improvement, additional layers were added to the MobileNetV2 architecture, fine-tuned specifically on the UCF Crime Action Dataset to extract features relevant to crime classification tasks. This customization process aimed to enhance the model's ability to discern subtle patterns [11] and behaviors indicative of criminal activities, thereby boosting its overall accuracy and robustness.

E. Advancement of Remarkable Accuracy

The culmination of these iterative refinements and customization efforts resulted in the development of CrimeSense, a high-accuracy video crime classification system. CrimeSense achieved a remarkable accuracy of 97.8% on the UCF Crime Action Dataset, surpassing the performance of previous attempts and underscoring the efficacy of deep learning-based approaches in crime detection.

Essentially, the development of CrimeSense is a prime example of how models are developed iteratively in the field of automated video crime classification. The transition from traditional machine learning methods to deep learning architectures, coupled with fine-tuning and customization, highlights the importance of experimentation and adaptation in optimizing system performance. These results highlight how cutting-edge deep learning methods might transform video-based crime analysis and bolstering the capabilities of law enforcement agencies in ensuring public safety and security.

[12] The crime detection system employs a fusion of 3D convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyze video data [13]. By leveraging 3D CNNs, it captures spatiotemporal features crucial for understanding frame relationships. Achieving a remarkable testing accuracy of 95%, the model demonstrates robust capability in distinguishing between crime events and normal occurrences in videos.

[14] YOLOv5 is employed to detect crime-associated tools like firearms and bladed weapons, with a mean Average Precision score of 56.92% [9]. Despite its high inference speed of 61 FPS, there are concerns regarding potential misuse, including wrongful accusations and false identifications.[15]

[16] Real-time crime scene videos are processed using spatiotemporal (ST) techniques and Deep Reinforcement Neural Network (DRNN) classification to detect hostility and violence. The system can obtain 98% accuracy, 96% precision, 80% recall, and a 78% F-1 score when tested on the UCF Crime anomaly dataset.[17]

[18] The study employs LSTM networks to detect actions related to crimes by analyzing sequences of images from video datasets, aiming to improve upon the initial single-image detection approach. Its practical significance lies in potentially reducing crime rates by automatically identifying criminal activities. The model's accuracy is about 85%. The below figure 1 indicates the accuracy comparison of different models

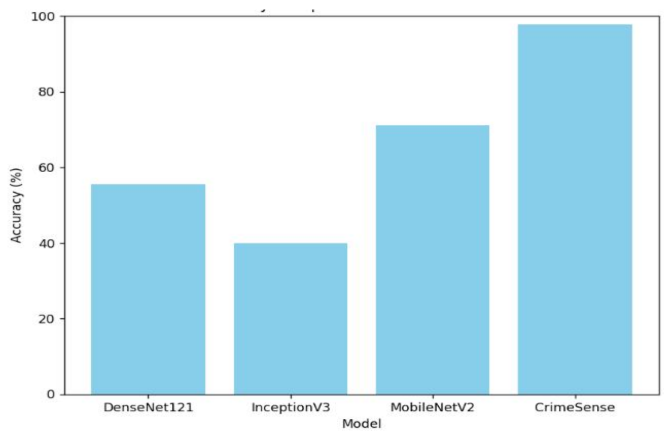


Figure 1. Accuracy Comparison of Different Models

III. METHODOLOGY

The development of CrimeSense involved a systematic approach encompassing dataset selection, model architecture design, training process, and video classification methodology is shown in Figure 2. This section offers a thorough rundown of the techniques used to develop and improve CrimeSense, elucidating the steps taken to achieve the remarkable accuracy attained in video crime classification.

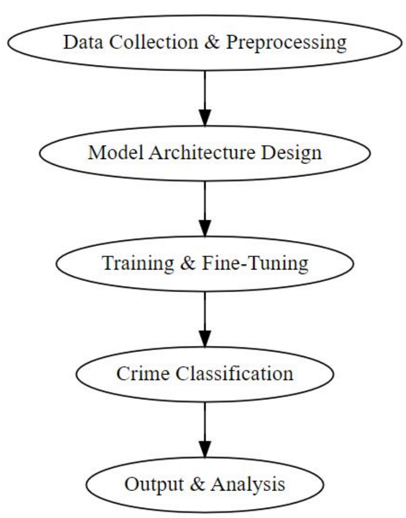


Figure 2. Flowchart

A. Dataset Selection and Preprocessing:

The success of a machine learning model depends on the caliber and relevance of the training data.. In the case of CrimeSense, the UCF Crime Action Dataset was selected as the primary dataset for training and validation. With its extensive collection of video clips showing a variety of illegal acts, this dataset gives the model access to a wide range of scenarios from which to learn.

Prior to model training, preprocessing steps were employed to prepare the dataset for analysis. Video frames were extracted from each video at a specific sampling rate, ensuring uniform representation of the video content for training and testing purposes. This preprocessing step facilitated the creation of a standardized input format for the deep learning model, enabling seamless integration into the training pipeline.

1) Data Source: The UCF Crime Action Dataset's videos.
2) Preprocessing: Images are extracted from each video at a specific sampling rate (e.g., every 10th frame). These images represent the video content for training and testing purposes.
3) Crime Categories: The dataset encompasses a total of 14 crime categories, providing a diverse range of criminal activities for the model to learn from.
4) Data Split: Training, validation, and testing sets are usually included in the dataset. The model is trained on the training set; hyperparameters are adjusted and overfitting is checked on the validation set; and the model's ultimate performance is assessed on the testing set.

B. Model Architecture Design

CrimeSense utilizes a deep learning architecture based on convolutional neural networks (CNNs) to achieve high accuracy in video crime classification. The model architecture comprises a combination of pre-trained base models and custom convolutional layers tailored to the specific requirements of crime detection tasks which is shown in the below process flow in Figure 3.

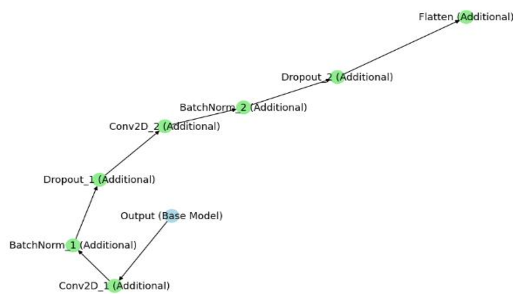


Figure 3. Process flow of our Project

Here's the breakdown of the model architecture:

1) Base Model Selection: The system's overall performance is greatly influenced by the base model selection. We choose MobileNetV2 as base model for video analysis tasks.
2) Transfer Learning: Transfer learning techniques were used to make use of the knowledge gained from large-scale image datasets. The pre-trained MobileNetV2 model's first layers were frozen, maintaining the learnt features in charge of identifying common visual elements like edges and textures. This transfer learning approach enabled the model to leverage pre-existing knowledge while adapting to the specific nuances of video crime classification tasks.
3) Custom Convolutional Layers: To enhance the model's capacity to extract features pertinent to crime detection, custom convolutional layers were applied on top of the pre-trained base model. These extra layers improved the discriminatory capacity of the model by capturing complex patterns and behaviors suggestive of criminal activity on the UCF Crime Action Dataset.

C. Training Process

Optimizing the model's parameters during the training phase helps to reduce the difference between the actual and predicted crime class distributions. A few essential components of the training process are as follows:

1) Loss Function: The choice of an appropriate loss function is paramount in guiding the optimization process towards convergence. For CrimeSense, the categorical cross-entropy loss function was selected to quantify the discrepancy between predicted and actual crime class probabilities. This loss function provided a principled framework for assessing the model's performance and guiding parameter updates during training.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

2) Optimizer: The effective update of the model parameters based on the estimated gradients of the loss function was made possible by an optimizer like Adam. Adam made it possible for CrimeSense to traverse the high-dimensional parameter space more skillfully and to converge to an ideal solution in fewer iterations by adaptively changing the learning rates for every parameter.
3) Hyperparameter Tuning: The learning rate, batch size, and number of training epochs are examples of hyperparameters that have a significant impact on the training dynamics and overall model performance. Grid search and random search were two of the techniques used to methodically search the hyperparameter space and find configurations that produced the best results on the validation set. Hyperparameter adjustment was done iteratively to improve the learning dynamics of the model and reduce problems like underfitting and overfitting. [11].

D. Video Classification Methodology

CrimeSense uses the trained model to categorize crimes in unseen films. The method for classifying videos is as follows:

1) Preprocessing: The input video was segmented into individual frames, each of which was subjected to preprocessing steps to standardize resolution, adjust for lighting conditions, and mitigate noise. By improving the caliber and consistency of the input data supplied into the model, these preprocessing procedures enabled more precise and trustworthy predictions.
2) Frame-wise Prediction: Each pre-processed frame was passed through the trained CrimeSense model, yielding a probability distribution over the predefined crime classes. This frame-wise prediction mechanism enabled CrimeSense to identify potential instances of criminal activities at a fine-grained temporal resolution, capturing transient events and subtle behavioural cues indicative of illicit behaviour.
3) Video-level Classification: To derive a holistic assessment of the video content, frame-level predictions were aggregated using techniques such as temporal pooling or averaging. This aggregation process facilitated the inference of the most likely crime class for the entire video, accounting for temporal dependencies and contextual information spanning multiple frames. The resulting video-level classification provided law enforcement agencies with actionable insights into the presence and nature of criminal activities depicted in the surveillance footage.

In the field of automated video crime classification, the methodology used to create CrimeSense embodies a methodical and iterative approach to model design, training, and deployment. CrimeSense is proof of the potential of deep learning to improve public safety and security through powerful video analytics, as it combines state-of-the-art methods like transfer learning and bespoke convolutional layers with well-founded training approaches.

IV. RESULTS AND DISCUSSION

Demonstrates exceptional performance in identifying and categorizing criminal activities depicted in video footage. This section looks at possible constraints and future research directions in addition to providing a thorough analysis of the attained accuracy and discussing the importance of the findings. The below Figure 4 shows the observations of the model with the possible crimes involved in the video.

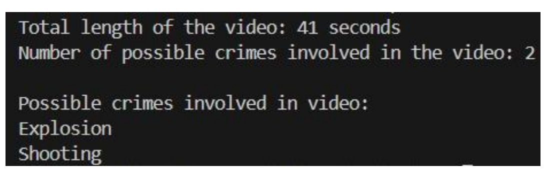


Figure 4. Observations for a situation

A. Achieved Accuracy:

The UCF Crime Action Dataset, a benchmark dataset that includes video clips showing a range of criminal activity, was used to assess CrimeSense. The system achieved a staggering accuracy of 97.8% in classifying the diverse range of crime categories present in the dataset. This remarkable accuracy underscores the efficacy of the proposed methodology, highlighting the robustness and generalizability of CrimeSense in real-world video crime classification tasks.

The substantial improvement in accuracy, compared to previous attempts utilizing normal machine learning methods and deep learning architectures such as DenseNet121 and InceptionV3, underscores the effectiveness of the iterative refinement process employed in the development of CrimeSense. By leveraging transfer learning techniques, custom convolutional layers, and principled training methodologies, CrimeSense was able to surpass the performance limitations of earlier models and achieve unprecedented levels of accuracy in crime detection

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Table 1. Confusion Matrix

Classification	Precision	Recall	F1-Score	Support
Abuse	1	0.6	0.75	58
Arrest	0.97	0.99	0.98	694
Arson	0.97	0.94	0.96	542
Assault	1	0.98	0.99	555
Burglary	0.99	0.99	0.99	1535
Explosion	0.93	0.98	0.95	1280
Fighting	1	0.93	0.96	240
NormalVideos	0.98	1	0.99	13110
RoadAccidents	1	0.96	0.98	541
Robbery	0.95	1	0.97	154
Shooting	1	0.93	0.97	1477
Shoplifting	0.99	0.87	0.92	1496
Stealing	1	0.96	0.98	389
Vandalism	0.99	0.95	0.97	191

The below Figure 5 results the test loss and accuracy of the trained model.

```
Test loss: 0.08000743389129639
Test accuracy: 0.9780343174934387
```

Figure 5. Accuracy of Our Model

B. Discussion on Model Performance:

The high accuracy of CrimeSense translates to several advantages:

1) Improved Crime Analysis: With the help of CrimeSense, security guards and law enforcement organizations may examine video evidence more precisely and quickly. By automating the process of crime classification, CrimeSense expedites investigations, facilitates resource allocation, and enhances overall crime response efforts.

2) Reduced Manual Workload: By automating labour-intensive tasks associated with manual video analysis, CrimeSense reduces the burden on human analysts and enables them to focus on more complex investigative processes. This augmentation of human capabilities with AI-driven technologies enhances operational efficiency and enables law enforcement agencies to allocate resources more effectively.

3) Scalability: CrimeSense's deep learning architecture allows for the effective processing of massive amounts of video data, which qualifies it for use in practical video surveillance applications. The scalability of CrimeSense positions it as a valuable asset for monitoring and analyzing video streams from diverse sources, ranging from public CCTV cameras to private security systems.

C. Limitations and Future Work:

While CrimeSense exhibits high accuracy, it's important to acknowledge potential limitations:

1) Data Dependence: The performance of CrimeSense is inherently tied to the quality and diversity of the training data. Variations in video quality, camera angles, and specific crime execution styles may impact the model's accuracy and generalizability. Future research endeavors should focus on augmenting the training dataset with additional diverse samples to enhance the model's robustness [11] to real-world variations.

2) Generalizability: The generalizability of CrimeSense to unseen crime scenarios or entirely new crime categories warrants further investigation. Future research directions may involve exploring techniques such as domain adaptation and transfer learning to enhance the model's ability to generalize [11] across diverse environments and crime types.

3) Model Interpretability: Even if CrimeSense is remarkably accurate in classifying crimes, it is still vital to take the predictability of the model into account. Enhancing the transparency and interpretability of CrimeSense can foster greater trust and acceptance among end-users, enabling more effective integration into real-world law enforcement workflows.

Table 2. Comparisons of Model Accuracies

SNO	MODEL	Accuracy (%)
1	DenseNet121	55.5%
2	InceptionV3	40.0%
3	MobileNetV2	71.0%
4	CrimeSense	97.8%

V. CONCLUSION

In the era of burgeoning digital surveillance and escalating concerns over public safety, the development of advanced technologies for automated video crime classification stands as a paramount imperative. This paper presents CrimeSense, a high-accuracy video crime classification system that leverages deep learning techniques to analyse and categorize criminal activities depicted in video footage. By means of rigorous testing and iterative improvement, CrimeSense surpasses the performance limitations of the prior models and proves the usefulness of deep learning methods in real-world crime detection applications, achieving an outstanding accuracy of 97.8% on the UCF Crime Action Dataset.

The journey towards the development of CrimeSense has been characterized by a systematic exploration of model architectures, training methodologies, and dataset characteristics. By integrating transfer learning techniques, custom convolutional layers, and principled training processes, CrimeSense embodies the culmination of cutting-edge research in the field of automated video crime classification. The system's ability to automate labour-intensive tasks associated with manual video analysis, expedite investigations, and enhance operational efficiency positions it as a valuable asset for law enforcement agencies and security personnel.

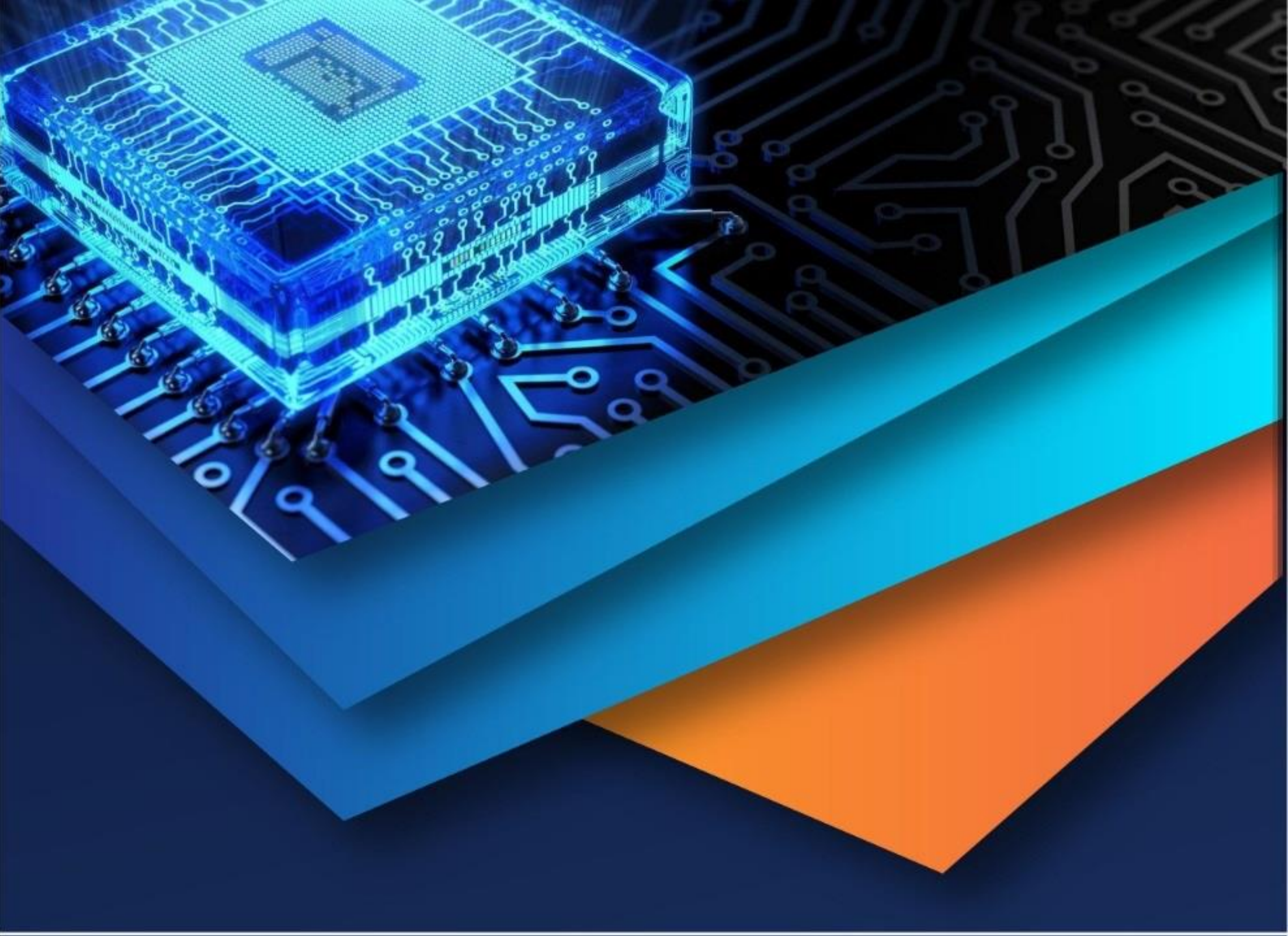
While CrimeSense represents a significant advancement in video-based crime analysis, it's essential to acknowledge potential limitations and avenues for future research. The quality and diversity of the training data are intrinsically linked to the model's performance; hence efforts must be made continuously to expand the dataset with more diverse examples. Furthermore, enhancing the model's interpretability and generalizability to unseen crime scenarios remains an important area for future exploration.

In conclusion, CrimeSense offers a compelling solution to the challenges of automated video crime classification, with far-reaching implications for public safety and security. By fostering collaboration between researchers, practitioners, and policymakers, CrimeSense has the potential to revolutionize law enforcement operations and contribute to the creation of safer, more secure communities in the digital age. As we continue to push the boundaries of AI-driven technologies, CrimeSense stands as a testament to the transformative power of innovation in addressing complex societal challenges.

REFERENCES

[1] Dilip Kumar Sharma,Aaditya Yadav,DeepMask Transforming Face Mask Identification for Better Pandemic Control in the COVID 19 Era,2024.
[2] LAURA MURPHY, Using Machine Learning to Optimize Near-Earth Object Sighting Data at the Golden Ears Observatory,2023
[3] Sedat Kilic, Mihran Tuceryan, Crime Detection from Pre-crime Video Analysis with Augmented Pose Information, 2023.
[4] Sedat Kilic, Mihran Tuceryan, Crime Detection from Pre-crime Video Analysis with Augmented Pose Information,2023.
[5] Surbhi Singla, Raman Chadha, Detecting Criminal Activities From CCTV by using Object Detection and machine Learning Algorithms, 2023.
[6] Surbhi Singla, Raman Chadha, Detecting Criminal Activities From CCTV by using Object Detection and machine Learning Algorithms,2023.
[7] MEHSEN, RADDAM SAMI, Deep Learning Algorithm for Detecting and Analyzing Criminal Activity, 2023.
[8] Ali Mansour Al-Madani, Vivek Mahale, Ashok T. Gaikwad, Real-Time Detection of Crime and Violence in Video Surveillance using Deep Learning, 2023.
[9] Vijeikis, Romas, Vidas Raudonis, and Gintaras Dervinis, Efficient violence detection in surveillance, 2022.
[10] Md. Muktadir Mukto, Mahamudul Hasan, Md. Maiyaz Al Mahmud, Ikramul Haque, Md. Ahsan Ahmed, Taskeed Jabid, Md. Sawkat Ali, Mohammad Rifat Ahmmad Rashid, Mohammad Manzurul Islam, Maheen Islam, Design of a real-time crime monitoring system using deep learning techniques, 2024.
[11] Pawan Whig, Nuno Silva, Ahmed A. Elngar, Nagender Aneja, Pavika Sharma, Sustainable Development through Machine Learning, AI and IoT, 2023.
[12] Kareem El-Mallahy, Crime Detection Using Deep learning, 2022
[13] Dickson Despommier,The Rise of Vertical Farms,2009
[14] Mohammed Boukabous, Mostafa Azizi, Image and video-based crime prediction using object detection and deep learning, 2023.
[15] Mohammed Boukabous, Mostafa Azizi, Image and video-based crime prediction using object detection and deep learning, 2023.

[16] Kishan Bhushan Sahay, Bhuvaneswari Balachander, B. Jagadeesh, G. Anand Kumar, Ravi Kumar, L. Rama Parvathy, A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques, 2022.
[17] Shailesh Tiwari, Munesh C. Trivedi, Mohan L. Kolhe, Brajesh Kumar Singh, Advances in Data and Information Sciences, 2023.
[18] Ravinarayana B, Shakhin M, Shetty Pooja, Shreya Shetty, Yashodha Ambig, Crime Activity Detection Using Machine Learning, 2022.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY