International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Enhancing CCTV Security: A Comprehensive Integration of Artificial Intelligence for Advanced Threat Detection and Prevention

Priyan S [1]***, Mr. K. Arunkumar*** [2]

[1]MSc Computer Science Rathinam College of Arts and Science Coimbatore, Priyans.mcs22@rathinam.in
[2]Department of Computer Science Rathinam College of Arts and Science Coimbatore

ABSTRACT :

Crime prediction in video surveillance systems is essential for preventing incidents and safeguarding assets. In our study, we introduce an innovative artificial intelligence approach for predicting and detecting Robbery Behavior Potential (RBP) in indoor camera footage. Our methodology involves three detection modules: head cover, crowd, and loitering detection, designed to facilitate timely interventions and deterrence of robberies. We train the first two modules by adapting the YOLOV5 model to our manually annotated dataset. Additionally, we introduce a novel loitering detection module based on the DeepSORT algorithm. A fuzzy inference machine is employed to translate expert knowledge into rules and make final predictions regarding potential robbery incidents. The complexity of our approach arises from factors such as variations in robber behavior, diverse camera angles, and low-resolution video imagery. Despite these challenges, we conducted experiments using real-world surveillance footage, achieving an F1-score of 0.537. To provide a benchmark for comparison, we set a threshold value for RBP and evaluated our method's performance against existing robbery detection techniques, resulting in a markedly improved F1-score of 0.607. We are confident that the adoption of our methodology could significantly mitigate the impact of robberies in surveillance control centers by enabling proactive prediction and prevention measures. Moreover, enhanced situational awareness among human operators can lead to more effective camera management and surveillance strategies.

Keywords: Artificial intelligence, Deep learning, Predictive Modelling, Surveillance Technology

1. Introduction :

Today, surveillance cameras are widely used in various places such as stores, banks, airports, and homes, to increase public safety and prevent the occurrence of crime. Alternatively, the time and place of the crime and specifically the wrongdoer can be achieved by analyzing these videos and aiming to identify the delinquent. Meanwhile, someone is needed behind the scene, watching the videos and noticing. The associate editor coordinating the review of this manuscript and approving it for publication was Varuna de Silva.

Whenever something anomaly is happening. However, due to very rare occurrence of an anomaly, the person becomes tired and if an anomaly happens, sometimes he cannot realize its occurrence. In other words, he loses the anomaly [1].

Furthermore, the anomaly-detection process is based on human common feeling which is learned during years. On the other hand, skill amount of the person for signs of crime occurrence understanding ability and the cost of employing him are other problems of non-automated crime prediction and detection systems which are based on watching surveillance videos.

To automate anomaly detection, some visual features must be extracted using machine learning and deep learning algorithms [2], [3]. For better performance of these algorithms, specific features for different anomaly classes [4] like vandalism [5], violence detection [6], and robbery [7] can be useful. Predicting the location and time of the crime reducing the destruction. On the other hand, security forces are also present on time such as an experiment, manufactured in Santa Cruz, California, where officers benefit from daily crime forecasts every morning. This forecasting navigates them to patrol determined regions. A Santa Cruz spokesperson declared that thirteen wrongdoers have been stopped in the determined areas during first the six month of experiment [8].

Due to paper [8], [9], and [10], some main symptoms prove that predictive policing is significant to be used for federal financing and security systems including: cost saving and crime reduction. Violent crimes are more dangerous because of their victimization probability and they increased by 20% due to Seattle Police Department (SPD) report during 2021 in Washington, USA [11].
According to statistics acquired from Federal Bureau of Investigations-Uniform Crime Reporting System (FBI-UCR), robbery is one of five common crimes in the United States [12].

The detection of robbery is one of the purposes of installing surveillance cameras in many places. Robbery is the crime of taking or attempting to grab any property by force, threat, or weapon [13] based on Oxford dictionary definition and differentiated from other forms of theft such as shoplifting, pickpocket, or burglary, by its intrinsically violent essence [14], [15]. While many lesser types of theft are punished as misdemeanors, robbery is always a felony in jurisdictions. Criminologists distinguish different types of robbery with regards to time and space of occurrence, armed or unarmed robberies, weapon types, and force amount. Therefore, one typical scissor is commercial robberies and street robberies [16].

Street robberies usually happen in poor crowded locations with no Closed-circuit televisions (CCTV). Commercial robberies occur in two ways: one where the offender enters the scene dressed up as a customer or conceal his face with normal covers like mask or helmets then suddenly out of the blue pulls a weapon and scares the employee. The other which offenders enter with force, typically in a group and probably conceal their face or head [17]. Both types of commercial robberies occurred in the indoor places which have customarily CCTVs so that detecting offenders or detection and even prediction of commercial robberies can be possible. Additionally, offenders who armed by weapon or knife usually threaten human with force. On the other hand, for offenders bearing any stick or be unarmed, a massive force is more probable [16], [17]. Hereupon, armed or unarmed commercial robberies force causes injury, pain, and even death.

Thus, predicting commercial robbery behavior by human, machine, or combination of these two, plays an important role in preventing its occurrence and its arisen dangers [18].

In general, there are some methods to automate detection or prediction of crimes based on extracting different crime scenarios and implementing them in different fields. But none of these methods have predicted the potential of robbery behavior. Therefore, there is a need to develop an algorithm for RBP prediction in video images. One could easily notice that, extracting the evidences and features in the surveillance videos is needed for prediction. To do this, the potential of robbery behavior in video images should be investigated. Scenarios of robbery occurrence, vary from one context to another [19] due to different conditions of each place selected for robbing and different cultures of countries. Therefore, robust feature extraction is not accompanied with certainty.

Despite the variety of robbery incidence scenarios and due to scenario-based approaches [20], [21], a common scenario with main points can be considered for commercial robbery videos. Specifically, one or some person choosing a poorly attended place who are usually covering their face or head by helmet, mask, glasses, or any garment to not be recognized and they are loitering to get an opportunity for showing their weapon, threat, or force. This scenario is completely matched with the knowledge of an expert person and definition of the first type of commercial robbery behavior [17], [22].

To implement a system based on this common scenario abstracted from different scenarios inferred from robbery videos, we consider common features found in most robbery cases under three modules including: head cover, crowd, and loitering detection. After extracting these features, for modules implementation, an inference machine is needed to conclude on the RBP. The conclusion process must be as competence as a human decision-making for potential derivation. Due to the ability of fuzzy set theory to mimic human inference [23], experience could be put in the form of fuzzy rules and according to fuzzy measurement, it facilitates the diagnosis and reasoning of a complex decision [24], [25].

Deep learning methods on the other hand, do not offer such adaptability and may not be able to deal with the nuances and variations of uncertain data well [25]. Owing to these reasons a fuzzy inference machine is proposed in this paper.

To sum up, main contributions of our paper are as below: 1. The proposed algorithm is based on a novel method which can predict RBP and prevent damages resulted by its occurrence in indoor places. To the best of our knowledge, this is the first work focusing on robbery behavior prediction and grounded on three main modules: Head cover, crowd, and loitering detection modules.

A dataset has been prepared for our system and annotated manually as two states: with or without head cover. For crowd counting, we sum the results of two states reported by head cover detection module. The method dominates the constraints of surveillance videos such as low resolution and single-camera videos.

The loitering point we have defined, is a novel definition for loitering calculation. A Deep Simple Online Real-time Tracking (DeepSORT) algorithm has used with respect to the tracking methods to calculate the amount of loitering for each person. By analyzing the obtained amount of loitering based on Euclidean distance calculation, a point has assigned to each one.

The key contribution of our algorithm is using a fuzzy inference machine with optimized rules, fuzzification, and defuzzification steps.

2. Related Works :

Anomalies are infrequent observations, events, or behaviors that stand out because they significantly deviate from normal patterns. Crime, as a form of anomaly, encompasses any behavior that strays from typical activities [2]. The widespread deployment of Closed-Circuit Television (CCTV) cameras can be attributed to the rising incidence of crimes in public spaces. Predicting crime involves detecting suspicious behaviors, a task that necessitates handling imperfect, ambiguous, and uncertain information [26].

Our proposed methodology focuses on predicting Robbery Behavior Potential (RBP) specifically within indoor environments. Robbery, a subtype of crime, necessitates

the detection of loitering, crowd dynamics, and head coverings. One key aspect of our approach is the development of a generic framework for RBP prediction, a feature not addressed in existing literature. In this section, we discuss several related studies pertinent to the detection or prediction of suspicious behaviors, crime, as well as research focusing on loitering and head cover detection.

Elhamod and Levine [27] introduced a semantics-based algorithm for recognizing suspicious behaviors, leveraging object tracking through blob matching with color histograms and spatial information. Ishikawa and Zin [18] proposed an automated system for detecting questionable pedestrians based on loitering behavior. Rajapakshe et al. [28] presented an E-police system incorporating video surveillance monitoring and crime prediction. Arroyo et al. [22] developed a real-time suspicious behavior detection system in shopping malls, utilizing image segmentation and tracking algorithms. Bouma et al. [29] focused on early detection of pickpocket behavior using pedestrian tracking and feature extraction. Selvi et al. [30] utilized enhanced CNN algorithms to detect actions like shooting and stealing. Roy [31] proposed a model for detecting snatch theft using Gaussian Mixture Models. Kaur et al. [32] developed algorithms for face mask detection, while Huang et al. [38] focused on helmet detection using improved YOLOV3.

Our proposed algorithm draws inspiration from these behavior detection systems, particularly those that employ scenario-based approaches. We leverage scenarios of robbery behavior to develop our method. Specifically, we build upon the work of Ishikawa and Zin [18] by utilizing the DeepSORT algorithm for person tracking and Euclidean distance for loitering detection. Additionally, we retrain YOLOV5 for head cover detection. A summary of these methods is provided in Table 1.

Importantly, none of the aforementioned studies have addressed RBP prediction with the aim of preventing its occurrence. Predicting robbery behavior involves assessing suspicious activities before any display of force or threat. Common robbery scenarios include targeting poorly attended stores, concealing one's identity, and loitering for an opportunity to commit the crime. Our approach involves detecting the number of individuals and their head coverings using YOLOV5, tracking individuals with DeepSORT, and employing a novel loitering calculation method based on Euclidean distance and defined displacement thresholds.

Fuzzy Logic, which mimics human reasoning with uncertainties, is utilized to infer RBP. This allows our system to make decisions based on imprecise inputs from the three detection modules, akin to human judgment. In the subsequent section, we detail our methodology and proposed approach.

3. Problem Formulation :

The increasing integration of artificial intelligence (AI) in security applications has paved the way for significant advancements in crime prevention. However, the accurate prediction of robbery behavior remains a challenging task. Traditional methods often lack precision and may result in false alarms, leading to suboptimal resource allocation for law enforcement agencies.

3.1 Nuanced Pattern Recognition

Existing systems may struggle to discern subtle nuances associated with pre-robbery behaviors in surveillance footage, requiring a more sophisticated approach for pattern recognition.

3.2 Adaptability to Evolving Patterns

The dynamic nature of criminal behavior necessitates a system that can adapt and learn continually, ensuring sustained accuracy over time as criminal tactics evolve.

3.3 Optimizing Resource Allocation

Minimizing false alarms is crucial for optimizing resource allocation in law enforcement. Developing a system that strikes a balance between sensitivity and specificity is essential for efficient utilization of resources.

3.4 Integration of Comprehensive Features

The incorporation of a diverse set of features, including motion patterns, object interactions, and facial expressions, is vital for creating a comprehensive predictive model capable of capturing multifaceted cues preceding a robbery attempt.

4. Proposed Methodology :

In this section, we delve into the detailed introduction of our proposed method for predicting RBP, as depicted in Fig. 1. We begin by outlining three primary blocks:

- Head cover detection module.
- Crowd detection module
- Loitering detection module.

Our dataset was meticulously curated by our team to align with the intended robbery scenario, facilitating the implementation of the head and crowd detection modules. Subsequently, the data underwent manual annotation and convolution to reduce its resolution. Through the customization of YOLOV5s via retraining, we have fully developed the first two modules. To track human movement, we utilize an Euclidean method to calculate the distance traveled, while employing the DeepSORT algorithm. By establishing individual thresholds, we assign the label of "loitering" to each person, thereby introducing our unique loitering detection module.

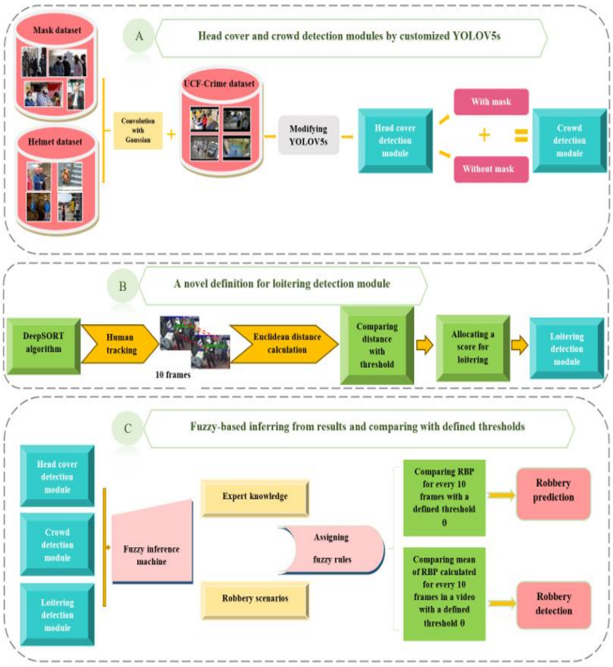
A
Head cover and crowd detection modules by customized YOLOV5s
Mask dataset
Helmet dataset
Convolution with Gaussian
+
UCF-Crime dataset
Modifying YOLOV5s
Head cover detection module
With mask
+
Without mask
=
Crowd detection module
B
A novel definition for loitering detection module
DeepSORT algorithm
Human tracking
10 frames
Euclidean distance calculation
Comparing distance with threshold
Allocating a score for loitering
Loitering detection module
C
Fuzzy-based inferring from results and comparing with defined thresholds
Head cover detection module
Crowd detection module
Loitering detection module
Fuzzy inference machine
Expert knowledge
Assigning fuzzy rules
Robbery scenarios
Comparing RBP for every 10 frames with a defined threshold θ
Comparing mean of RBP calculated for every 10 frames in a video with a defined threshold θ
Robbery prediction
Robbery detection

Figure 1 – Block diagram

4.1 Head Cover Detection

In our proposed method, head cover refers to any item such as a hat, helmet, mask, glasses, or clothing that conceals a person's head and face, hindering their identification. Robbers often opt for head coverings to avoid recognition during robberies. We propose a method capable of detecting human heads in stores, with or without head coverings, in low-resolution single frames, and labeling them as masked or unmasked.

Our proposed algorithm utilizes a deep learning method for the head cover module. You Only Look Once (YOLO) is a deep learning algorithm used to detect objects by treating the entire image as a regression problem. YOLO divides the image into grids to extract global information and detect objects. Given that surveillance video resolutions are typically low, the head cover detection module must overcome this challenge. YOLO algorithms can be retrained using low-resolution images to achieve accurate detection results. YOLOv5, an improved version of previous YOLO models, offers higher processing speed and better capability in detecting small objects, making it suitable for our purposes.

We choose YOLOv5 for our proposed method and retrain it using prepared low-resolution images. YOLOv5 adapts the width and depth of the backbone network, ensuring high human and object detection accuracy. We opt for the YOLOv5s version, known for its simplicity, small size, and speed.

4.2 Crowd Detection

The presence of people in the store, whether with or without head coverings, is indicative of crowd density. Thus, the crowd detection module is essentially derived from the head cover detection module's results. By determining the number of detected human heads, we can assess the level of crowding in the environment. A lower attendance rate poses a higher risk of robbery, whereas increased human presence reduces the likelihood of robbery, as defined in Equations 1 and 2.

$$C = \left\{0.1, \frac{2}{\{N\}}\right\} \qquad (1)$$

$$C^s = \begin{cases} 100 \times MinC & \text{N} = 1 \text{ or N} \geq 7 \\ 100 \times MaxC & 1 < N < 7 \end{cases} \qquad (2)$$

Let C be crowd corresponding set, N be the number of people and Cs be crowd score. As a result, Cs is between 0 100 and showing the number of people present in the stores. Based on crowd score we defined, the effect of the presence of people can be considered on RBP.

4.3 Loitering Detection

Individuals intending to make purchases in a store typically browse for items before heading to the cash desk to complete their transactions. However, a robber may exploit this opportunity to plan a robbery, exhibiting suspicious behavior such as lingering around the counter area. This increased loitering activity by the robber around the counter is detected through surveillance cameras focused on the counter area.

To evaluate the degree of loitering, each person is tracked using the DeepSORT algorithm. The total distance traveled by each person is calculated over successive frames, with a sliding window approach applied to the video data. Specifically, the video is divided into overlapping snippets with a fixed number of frames, termed a "snippet," and each snippet is analyzed for loitering activity. The distance traveled by individuals is monitored over each snippet, with updates made at regular intervals.

It's important to note that the cameras used to collect the dataset for this study were not calibrated, resulting in distance calculations being scale-dependent. However, due to the minimal movement within each 10-frame interval, the distance and displacement of individuals are considered equal in our algorithm.

the process of snippet allocation and loitering calculation. The video's length (L) is divided into overlapping snippets (S) with a fixed step size of 10 frames. If an individual exhibits loitering behavior around the counter, their total traveled distance exceeds a predetermined threshold (θn). Based on the extent of movement, each person is assigned a loitering score between 0 and 100.

Equation 3 outlines the division of the video into steps of 10 frames each, facilitating the calculation of the distance traveled during each step. These distances are then aggregated within each snippet, allowing for the detection of loitering behavior over successive intervals.

$$n = [\frac{L}{10}] \qquad (3)$$

Eq. 4 elucidates Euclidean distance calculation and aggregation of them for every snippet.

$$\begin{cases} i = 10j & \text{j} = 0 : (\text{n} - 1) \\ d_j = \sqrt{(C_{x(i+10)} - C_{x_i})^2 + (C_{y(i+10)} - C_{y_i})^2} \\ D^m = \sum_{i=m}^{50+m} d_j \quad \text{m} = 0 : (\text{j} - 50) \end{cases} \qquad (4)$$

where dj shows displacement of a human from ith frame to i+10 during one snippet. Besides, t is the number of steps during one snippet and it is equal to 50 because 50 × 10 = 500 frames. Furthermore Cx,y shows position of head part for each human and their displacement during 10 frames is shown with i and i+10. di is Euclidean distance calculated for every 10 frames and Dm is aggregation of them for every snippets consisting 500 frames (50 × 10). After calculation of Dm for one snippet, one step slides during the video and the aggre gation is calculated again. This exertion is for consideration of the human mannerchanges during the video.

4.4 Fuzzy Inference Machine

Fuzzy logic theory is grounded on the concept of fuzzy sets, where the degree of membership function delineates the connection between a member and the set, allowing for intermediate states of membership. This theory enables the quantification of all modules through intermediate values using appropriate membership rules and functions to assess the potential of robbery [24].

There are two primary types of fuzzy inference methods:

- Mamdani approach, which employs linguistic fuzzy modeling.
- Takagi-Sugeno-Kang (TS) approach, based on precise fuzzy modeling.

The Mamdani approach prioritizes interpretability over accuracy, while TS emphasizes accuracy with lower interpretability. Given the interpretability of features extracted by the three modules and the conceptual interpretability of the Mamdani approach, it is the suitable inference machine for calculating the potential of robbery.

The Mamdani fuzzy approach involves four steps: fuzzification, inference, composition, and defuzzification. Fuzzification compares input variables with membership functions (MF) to assign corresponding membership values to each element. In the inference step, membership values are combined based on the premise step to obtain the fulfillment command. Composition produces fuzzy or crisp consequences, and defuzzification aggregates consequences using MFs to generate a crisp output. Triangular, trapezoidal, and Gaussian functions are commonly used MFs in Mamdani fuzzy inference systems. Defuzzification converts the fuzzified output into a crisp value.

In this algorithm, the modules of head cover, crowd, and loitering detection are regarded as input variables, while the potential of robbery behavior serves as the output variable. A fuzzy inference machine utilizing Mamdani rules has been devised for this purpose. Triangular MF is chosen for its simplicity, computational efficiency, and incorporation of expert knowledge, requiring three parameters (a, b, and c) to define triangular MF, as depicted in Eq. 5.

$$\mu_j = max\left(min\left(\frac{x_i - a}{b - a}, \frac{c - x_i}{c - b}\right), 0\right), \quad (a < b < c) \quad (5)$$

Parameters a,b andcarecoordinates of three corners from the triangular MF and they are acquired from experts’ knowledge for all of our detection modules and RBP prediction and detection system.

5. Results and Discussion :

In our empirical evaluation, we focus on two scenarios: predicting robbery potential and detecting actual robbery incidents. To validate our proposed system, we assemble a suitable dataset for retraining deep learning algorithms, specifically tailored for extracting individuals, identifying their head cover, and detecting loitering behavior.

For the prediction phase, we select 45 videos from the Robbery-UCF-Crime dataset [1]. These videos capture instances where a potential robber is observed loitering, indicating an imminent robbery attempt. It's important to note that prediction involves identifying behaviors that have not yet occurred. In the context of robbery, prediction is feasible until the moment the robber reveals a weapon, makes threats, or uses force. Individuals with head cover may loiter in low-crowded areas of shopping malls, waiting for an opportunity to execute a robbery.

On the other hand, we utilize 70 videos from the Robbery-UCF-Crime dataset for the robbery detection algorithm. Detection involves identifying the behavior after it has occurred, such as when the robber reveals a weapon, makes threats, or employs force, and possibly leaves the scene after committing the robbery. These videos are carefully selected to ensure proper camera angles, particularly focusing on visibility around cash desks where robbery incidents commonly occur.

To calculate the potential for robbery in each scene, we employ a fuzzy inference machine. This machine simulates the decision-making process of the human brain by assigning a potential score based on the outputs of the three modules: head cover detection, crowd detection, and loitering detection. This section presents the experimental results to evaluate the performance of our proposed algorithm in both predicting and detecting robbery incidents.

5.1 Data Preparation

To enhance the performance of YOLOv5s for head cover detection in low-resolution images, we curated an image dataset sourced from three distinct groups: video image sequences extracted from anomaly folders of the UCF-Crime dataset (excluding the robbery folder) [1], the Bikes Helmets Dataset, and the Mask Dataset.

The dataset was meticulously assembled to include images captured from CCTV camera angles with significant variation in human positions and backgrounds, ensuring diverse and representative samples. In total, the dataset comprises 7621 images, with 5129 from the Bikes Helmets Dataset, 254 from the Mask Dataset, and 2238 from the UCF-Crime dataset (excluding robbery videos), which were converted into image sequences.

These images were then split into training, validation, and test sets with an approximate ratio of 7:2:1, respectively. Detailed information about the dataset is provided in Table 2.

For images extracted from the UCF-Crime dataset, manual selection was conducted to ensure significant changes in backgrounds or human positions. Subsequently, the selected images were meticulously annotated using the Computer Vision Annotation Tool (CVAT), a precise tool for localizing bounding boxes and generating high-quality annotations. The annotations focused on identifying the head part of humans, with each image labeled as either "masked" for individuals wearing head cover or "no-mask" for those without.

The annotations were saved in ".txt" format, containing information about the class, center coordinates (xc and yc), width, and height of each bounding box. These coordinates were normalized to the dimensions of the image using the following equation:

This normalization process ensures consistency and compatibility across all annotated images and for making all input images uniform.

$$\begin{cases} W = \left(\dfrac{x_{max} - x_{min}}{w}\right) \\ H = \left(\dfrac{y_{max} - y_{min}}{h}\right) \\ X = \left(\dfrac{x_{max} + x_{min}}{2 \cdot w}\right) \\ Y = \left(\dfrac{y_{max} + y_{min}}{2 \cdot h}\right) \end{cases} \qquad (6)$$

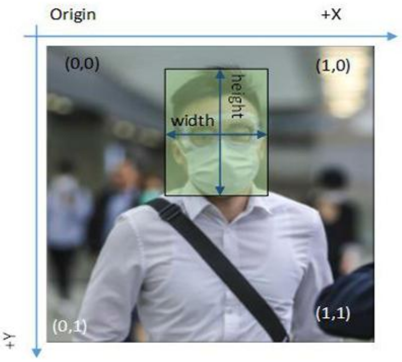

Figure 2 – Coordinates Position belongs to bounding box of human head

Figure 3 – Sample image from dataset

5.2 Fuzzy Inference Machine

The most suitable inference machine for integrating information from different modules and predicting the potential of robbery behavior is the fuzzy inference machine. This is because it can evaluate all modules using intermediate values with appropriate membership rules and functions for computing RBP [53]. The input variables for the fuzzy inference machine are derived from three modules: head cover detection, human detection, and loitering computation, while the output variable is the potential of robbery behavior.

To determine the potential of robbery behavior, a fuzzy inference machine with Mamdani rules and triangular membership functions for the inputs is employed. The inputs are categorized into linguistic variables with specific threshold values. For the head cover detection module, three threshold values ("Low", "Medium", and "High") are defined based on the number of head cover types detected by a person. Similarly, for human detection and loitering computation, five threshold values ("Very-Low", "Low", "Medium", "High", and "Very-High") are established to quantify the quantity of individuals and the degree of loitering, respectively.

In total, 75 rules (5 × 5 × 3) are formulated in the fuzzy inference machine to model the relationships between the values obtained from the modules. These rules are devised to capture the impact of each module on increasing or decreasing the robbery potential, drawing on the expertise of individuals familiar with surveillance videos and robbery behavior. An excerpt of the rules is presented in Table 5, highlighting their diverse nature across different intervals of robbery potential.

The defuzzification method employed in this study is the centroid strategy, which provides a comprehensive representation of the meaning derived from the input modules by blending all contributing rules. This method ensures a balanced consideration of the inputs and their respective contributions to determining the potential of robbery behavior.

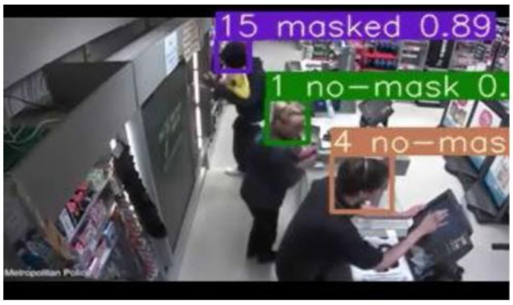

Figure 4 - Head cover Detection

5.3 Discussion :

Your approach to predicting and detecting commercial robbery is quite comprehensive and addresses several important challenges. Let's summarize the key points: You have chosen to focus specifically on commercial robbery, recognizing the importance of predicting such incidents to prevent financial and life-threatening events. Your methodology involves three main modules: crowd detection, head cover detection, and loitering detection, along with a fuzzy inference machine for calculating robbery potential. Gathering suitable data for training your detection modules posed challenges due to variations in camera viewpoints, image resolutions, and the need to capture various types of head coverings. Both head cover and human detection modules are implemented simultaneously using a dataset tailored to the specifics of CCTV camera viewpoints and low-resolution images. Head coverings, including masks and helmets, are crucial indicators of suspicious behavior, especially in the context of COVID-19. They increase the potential for robbery by obscuring the identity of perpetrators. Your loitering detection module utilizes the DEEP SORT algorithm for tracking individuals, with the goal of identifying suspicious behavior based on movement patterns. Unlike previous crime detection methods that focus on generic features, your approach is scenario-based, specifically tailored to the characteristics of robbery behavior. This approach offers flexibility for different cultural contexts and manifestations of robbery. The fuzzy inference machine is used to infer robbery potential, with rules optimized based on expert input. This allows for fine-tuning of the system based on real-world observations. You acknowledge limitations in the loitering detection module, particularly in cases of low-resolution images and instances of human overlap, which can lead to errors in tracking and identification. Overall, your approach demonstrates a thoughtful and systematic methodology for predicting and detecting commercial robbery in surveillance videos, addressing key challenges and leveraging advanced techniques such as deep learning and fuzzy inference.

6. Conclusion :

This research presents an approach for predicting Robbery Behavior Potential (RBP) in video surveillance images, aiming to address the challenges posed by CCTV videos, such as varying robbery scenarios, diverse camera angles, and low image resolution. The objective is to enable timely intervention to prevent or mitigate robbery incidents observable from surveillance footage. This study is motivated by the lack of prior work on RBP prediction despite the importance of preventing robbery occurrences.

Common robbery scenarios are identified through expert insights and analysis of CCTV footage, allowing for the extraction of key features. A deep learning-based approach, coupled with fuzzy inference, is proposed to assess the potential for robbery. The method involves retraining the YOLOv5 algorithm using a dataset containing images of individuals with and without head cover, facilitating efficient detection of crowd and head cover. The loitering module is implemented using a defined methodology that calculates individuals' Euclidean traveled distances with the DeepSORT algorithm.

A fuzzy inference machine is employed to infer the robbery potential for every 10 frames and aggregate the results for each snippet based on the outputs of the three modules. The proposed method is evaluated on the Robbery folder of the UCF-Crime dataset, achieving an F1-score of 0.537, indicating its capability to predict robbery potential accurately in over half of the videos.

Furthermore, the problem is transformed from prediction to detection of robbery, enabling comparison with existing literature on anomaly detection, particularly robbery detection, using the UCF-Crime dataset. The detection method achieves an F1-score of 0.607, outperforming other methods and demonstrating the effectiveness of the proposed scenario-based system in detecting and predicting robbery behavior.

The proposed approach can be applied in various settings equipped with surveillance cameras to prevent robbery crimes without the need for constant human monitoring of video feeds. Future enhancements could focus on improving the accuracy of loitering detection, potentially increasing the F1-score. Additionally, efforts may be directed toward developing an improved tracking algorithm for low-resolution video images, possibly by retraining the YOLOv5 algorithm specifically for low-resolution human detection.

7. Future Work :

Moving forward, several avenues for future research emerge from this project. Firstly, conducting further experimentation and validation on larger and more diverse datasets can provide additional insights into the robustness and generalizability of the proposed system. Additionally, exploring the integration of multimodal data sources, such as audio and text, could enhance the system's ability to detect and predict robbery behavior more accurately. Furthermore, investigating the potential use of real-time data streams and adaptive learning algorithms can enable the system to adapt more effectively to dynamic environments and evolving criminal tactics. Moreover, collaborating with law enforcement agencies and stakeholders to deploy the system in real-world settings and evaluating its impact on crime rates and community safety would be invaluable. Finally, continuous research and development efforts are needed to stay abreast of technological advancements and ensure the ongoing optimization and refinement of the AI-driven predictive model. By pursuing these avenues, future studies can build upon the foundation laid by this project and further advance the field of intelligent security systems for proactive crime prevention.

REFERENCES :

[1] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6479–6488.

[2] J. James P. Suarez and P. C. Naval Jr., "A survey on deep learning techniques for video anomaly detection," 2020, arXiv:2009.14146.

[3] T. Mei and C. Zhang, "Deep learning for intelligent video analysis," in Proc. 25th ACM Int. Conf. Multimedia, Oct. 2017, pp. 1955–1956.

[4] H. Yan, X. Liu, and R. Hong, "Image classification via fusing the latent deep CNN feature," in Proc. Int. Conf. Internet Multimedia Comput. Service, Aug. 2016, pp. 110–113.

[5] M. Ghazal, C. Vazquez, and A. Amer, "Real-time automatic detection of vandalism behavior in video sequences," in Proc. IEEE Int. Conf. Syst., Man Cybern., Oct. 2007, pp. 1056–1060.

[6] I. P. Febin, K. Jayasree, and P. T. Joy, "Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm," Pattern Anal. Appl., vol. 23, no. 2, pp. 611–623, May 2020.

[7] W. Lao, J. Han, and P. De With, "Automatic video-based human motion analyzer for consumer surveillance system," IEEE Trans. Consum. Electron., vol. 55, no. 2, pp. 591–598, May 2009.

[8] A. G. Ferguson, "Predictive policing and reasonable suspicion," Emory Law J., vol. 62, no. 2, p. 259, 2012.

[9] C. Beck and C. McCue, "Predictive policing: What can we learn from Wal-Mart and Amazon about fighting crime in a recession?" Police Chief, vol. 76, no. 11, p. 18, 2009

[10] K.J. Bowers and S.D. Johnson, "Who commits near repeats? A test of the boost explanation," Western Criminol. Rev., vol. 5, no. 3, pp. 12–24, 2004.

[11] Seattle Police Department, "SPD 2021 year-end crime report," Seattle, WA, USA, 2021. [Online]. Available: https://www.seattle.gov/documents/Departments/Police/Reports/2021_SPD_CRIME_REPORT_FINAL.pdf

[12] FBI, "Crime in the U.S. 2019," 2019. [Online]. Available: https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/robbery

[13] B. Fawei, J.Z. Pan, M. Kollingbaum, and A.Z. Wyner, "A semi-automated ontology construction for legal question answering," New Gener. Comput., vol. 37, no. 4, pp. 453–478, Dec. 2019.

[14] R. Thompson, "Understanding theft from the person and robbery of personal property victimisation trends in England and Wales," Nottingham Trent Univ., Nottingham, U.K., Tech. Rep. 2010/11, 2014.

[15] P. J. Cook, "Robbery violence," J. Criminal Law Criminol., vol. 78, no. 2, pp. 357–376, 1987.

[16] J. D. McCluskey, "A comparison of Robbers' use of physical coercion in commercial and street robberies," Crime Delinquency, vol. 59, no. 3, pp. 419–442, Apr. 2013.

[17] D. F. Luckenbill, "Patterns of force in robbery," Deviant Behav., vol. 1, nos. 3–4, pp. 361–378, Apr. 1980.

[18] T. Ishikawa and T. T. Zin, "A study on detection of suspicious persons for intelligent monitoring system," in Proc. Int. Conf. Big Data Anal. Deep Learn. Appl., Singapore: Springer, 2018, pp. 292–301.

[19] J. R. Medel and A. Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," 2016, arXiv:1612.00390.

[20] M. Shah, O. Javed, and K. Shafique, "Automated visual surveillance in realistic scenarios," IEEE Multimedia Mag., vol. 14, no. 1, pp. 30–39, Jan. 2007.

[21] A. Biswas, S. C. Ria, Z. Ferdous, and S. N. Chowdhury, "Suspicious human-movement detection," Ph.D. dissertation, Dept. Comput. Sci. Eng., BRAC Univ., Dhaka, Bangladesh, 2017.

[22] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazán, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," Expert Syst. Appl., vol. 42, no. 21, pp. 7991–8005, Nov. 2015.

[23] R. Nawaratne, D. Alahakoon, D. De Silva, and X. Yu, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance," IEEE Trans. Ind. Informat., vol. 16, no. 1, pp. 393–402, Jan. 2020.

[24] F. Wu, G. Jin, M. Gao, Z. He, and Y. Yang, "Helmet detection based on improved YOLO V3 deep model," in Proc. IEEE 16th Int. Conf. Netw. Sens. Control (ICNSC), May 2019, pp. 363–368.

[25] Y. Liu, X.-K. Wang, W.-H. Hou, H. Liu, and J.-Q. Wang, "A novel hybrid model combining a fuzzy inference system and a deep learning method for short-term traffic flow prediction," Knowl.-Based Syst., vol. 255, Nov. 2022, Art. no. 109760.

[26] V. Vaidehi, S. Monica, S. M. S. Safeer, M. Deepika, and S. Sangeetha, "A prediction system based on fuzzy logic," in Proc. World Congr. Eng. Comput. Sci., 2008, pp. 1–6.

[27] M. Elhamod and M. D. Levine, "Automated real-time detection of potentially suspicious behavior in public transport areas," IEEE Trans. Intell. Transp. Syst., vol. 14, no. 2, pp. 688–699, Jun. 2013.

[28] 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," in Proc. Int. Conf. Advancements Comput. (ICAC), Dec. 2019, pp. 310–316.

[29] H. Bouma, J. Baan, G. J. Burghouts, P. T. Eendebak, J. R. van Huis, J. Dijk, and J. H. C. van Rest, "Automatic detection of suspicious behavior of pickpockets with track-based features in a shopping mall," Proc. SPIE, vol. 9253, pp. 112–120, Oct. 2014.

[30] E. Selvi, M. Adimoolam, G. Karthi, K. Thinakaran, N. M. Balamurugan, R. Kannadasan, C. Wechtaisong, and A. A. Khan, "Suspicious actions detection system using enhanced CNN and surveillance video," Electronics, vol. 11, no. 24, p. 4210, Dec. 2022.

[31] D. Roy and K. M. C., "Snatch theft detection in unconstrained surveillance videos using action attribute modelling," Pattern Recognit. Lett., vol. 108, pp. 56–61, Jun. 2018.

[32] G. Kaur, R. Sinha, P. K. Tiwari, S. K. Yadav, P. Pandey, R. Raj, A. Vashisth, and M. Rakhra, "Face mask recognition system using CNN model," Neurosci. Informat., vol. 2, no. 3, Sep. 2022, Art. no. 100035.

[33] J. Ieamsaard, S. N. Charoensook, and S. Yammen, "Deep learning-based face mask detection using YOLOV5," in Proc. 9th Int. Electr. Eng. Congr. (iEECON), Mar. 2021, pp. 428–431.

[34] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," Sustain. Cities Soc., vol. 66, Mar. 2021, Art. no. 102692.

[35] S. Sethi, M. Kathuria, and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of coronavirus spread," J. Biomed. Informat., vol. 120, Aug. 2021, Art. no. 103848.

[36] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of InceptionV3," in Proc. Int. Conf. Big Data Anal. Cham, Switzerland: Springer, 2020, pp. 81–90.

[37] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," Multimedia Tools Appl., vol. 80, no. 13, pp. 19753–19768, May 2021.

[38] L. Huang, Q. Fu, M. He, D. Jiang, and Z. Hao, "Detection algorithm of safety helmet wearing based on deep learning," Concurrency Comput., Pract. Exper., vol. 33, no. 13, p. e6234, 2021.

[39] F. Zhou, H. Zhao, and Z. Nie, "Safety helmet detection based on YOLOv5," in Proc. IEEE Int. Conf. Power Electron., Comput. Appl. (ICPECA), Jan. 2021, pp. 6–11.

[40] T. Choudhury, A. Aggarwal, and R. Tomar, "A deep learning approach to helmet detection for road safety," J. Sci. Ind. Res., vol. 79, no. 6, pp. 509–512, 2020.

[41] R. Nayak, M. M. Behera, V. Girish, U. C. Pati, and S. K. Das, "Deep learning based loitering detection system using multi-camera video surveillance network," in Proc. IEEE Int. Symp. Smart Electron. Syst. (iSES) (Formerly iNiS), Dec. 2019, pp. 215–220.

[42] Y. Nam, "Loitering detection using an associating pedestrian tracker in crowded scenes," Multimedia Tools Appl., vol. 74, no. 9, pp. 2939–2961, May 2015.

[43] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[44] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," Comput. Electron. Agricult., vol. 157, pp. 417–426, Feb. 2019.

[45] S. Pouyan, M. Charmi, A. Azarpeyvand, and H. Hassanpoor, "Significantly improving human detection in low-resolution images by retraining YOLOv3," in Proc. 26th Int. Comput. Conf., Comput. Soc. Iran (CSICC), Mar. 2021, pp. 1–6.

[46] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," Electronics, vol. 9, no. 3, p. 537, Mar. 2020.

[47] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, arXiv:2004.10934.

[48] D. Luvizon, H. Tabia, and D. Picard, "SSP-Net: Scalable sequential pyramid networks for real-time 3D human pose regression," 2020, arXiv:2009.01998.

[49] Q. Z. Li and M. H. Wang, "Development and prospect of real time fruit grading technique based on computer vision," Trans. Chin. Soc. Agricult. Machinery, vol. 30, no. 6, pp. 1–7, 1999.

[50] W. Jia, S. Xu, Z. Liang, Y. Zhao, H. Min, S. Li, and Y. Yu, "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector," IET Image Process., vol. 15, no. 14, pp.