Aman Yadav, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 12, Issue 12, December 2022, pp. 74-79

RESEARCH ARTICLE OPEN ACCESS

Surveillance System using Computer Vision

Aman Yadav
U.G. Student, Department of Computer Science & Engineering
J.S.S. Academy of Technical Education
Noida, India

Surekha M
Assistant Professor, Department of Computer Science & Engineering
J.S.S. Academy of Technical Education
Noida, India

Date of Submission: 04-12-2022 Date of Acceptance: 15-12-2022

***Abstract*—**This paper proposes a Fraudulent Detection System, for use in financial institutions such as banks and ATMs to prevent potential robbery. Law enforcement agencies devote vast monetary and non-monetary resources like guards, officers, CCTV, and deploy police mobiles for patrol, to reach the crime scene in the minimum time possible. Still, the immediate arrest of offenders and protection of citizens are not assured. Therefore, we attempt to minimize the response time for the prevention of crime using techniques such - Object detection, pose detection, and alert system. We will use OpenCV, deep learning, and WhatsApp for the same.

Keywords— Object Detection, Pose detection, YOLO, Caffe Model

I. INTRODUCTION

Looking at various case studies of bank robbery and statements given by the offenders, we got to know that bank robbery mostly tends to happen in cities or towns which are commercial hubs and whereseveral retail shops are located. As urban areas offer maximum money deposited and a crowd of people in which they can blend in and hold hostages. Closed-circuit television systems (CCTV) are widely used all over the world including in banks and ATMs. Video Surveillance is mainly monitored by its operator to find any unusual activity. The major issue that persists in the existing bank security state is that it requires people in charge to manually trigger the security alarm [2].

The offenders are able to surpass all the conventional security systems used such as burglar alarms, access control systems, and CCTVs. Moreover, a burglar alarm alerts the offender, who may take extreme action like shooting hostages and may cause harm before law enforcement officials reach the site of the crime. So, there is a need for a smart burglar alarm that detects the offender before the crime takes place entirely [3].

Identifying certain parameters like a helmet, mask, or gun that cause suspicion, the system will send an immediate alert analysing the situation through the surveillance cameras present in banks, without endangering human life.

II. LITERATURE REVIEW

A large amount has been performed for object detection and pose detection for preventing the situations like bank robberies. However, there are limitations associated with those systems in the real world. Previously people have used techniques such as sliding window object detection, R CNN, Fast R CNN, and Faster R CNN. The CNN is used to predict various class probabilities and the bounding boxes simultaneously. It is not capable of real-time detection using its two-step architecture.

Deep learning has improved its methods to successfully recognize and detect criminal activity [4]. Artificial intelligence based on Deep Convolutional Neural Networks (CNN) with a tracking algorithm is used to detect the human body's position inside the surveillance video [5].

The work reported in this paper is far more different and improved to tackle the problem statement in front of us. In comparison to all of the previous works, it uses the You only look once (YOLO) algorithm to detect the real-time object rather than Faster R-CNN architecture. YOLO architecture has great merit over Faster R-CNN in detection speed as the frames per second (FPS) was more than eight times than that of Faster R-CNN architecture [6].

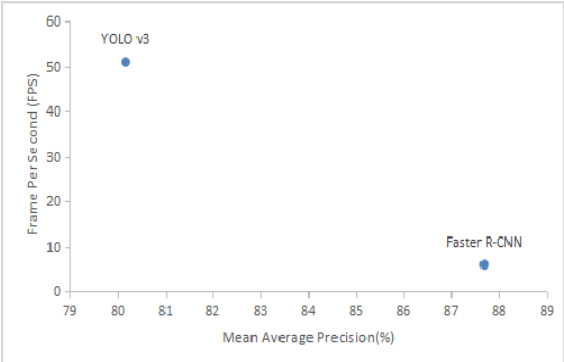

Fig 1. YOLO v3 Vs Faster R-CNN.

III. PROBLEM DEFINITION

Our system can be described as an autonomous system that can scan the frames per second from a live feed CCTV video footage, using the OpenCV library and identify objects if it lies in these three classes - helmet, gun, and/or mask. The detection of these three classes is performed using a manually trained dataset by the YOLO algorithm. Pose detection is also implemented by Deep Learning with the help of Caffe model to detect the patterns from environments and identify if there is a possible threat to the bank.

If we can identify potential robbery, then we can decrease the response time for alerting police officials using WhatsApp, Ideally leading to the arrest of the criminals before inflicting any harm to people.

IV. PROPOSED SYSTEM METHODOLOGY

A. OpenCV

It is the library containing programming algorithms and functions which are used for image processing. It is used to solve a number of real-time problems using its optimized algorithm. It's functions can be used to identify faces, objects, observe camera movement, and develop 3D models of different objects [7] [8].

B. YOLO (You only look once)

YOLO is a state-of-the-art object detection algorithm that has become the main method of detecting objects in the field of computer vision [9].

YOLO is highly efficient because of its speed, accuracy & learning capabilities YOLO algorithm works using residual blocks, bounding box regression, and intersection union.

YOLO algorithms can be applied in the autonomous driving vehicle, wildlife & various security fields [10].

In this paper, we provide a demonstration of the usage of the newest YOLOv3 algorithm for the detection of bank intrusion. We will train the network for 3 object classes (gun, mask, and helmet) and demonstrate the effectiveness of the approach in a variety of conditions by training models by providing thousands of images.

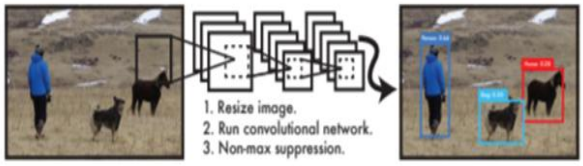

Fig 2. Image processed by resizing, running it through convolution layer and then threshold the result using model's confidence.

C. CaffeModel

Caffe Model is used for Real-time pose detection. It provides an adjustable framework for state-of-the-art algorithms for deep learning and various reference models. It supports various large-scale projects and is maintained by Berkeley Vision and Learning Center (BVLC) with help of the working community on Github platform [11] [12].

In Image processing, pose detection refers to understanding the movement of joints and extracting features from a processed and informative dataset. For understanding, if robbery is taking place, we can also analyse the pose of people present inside the compound [13].

If there are people present inside the bank with their hands up then it could be a possible sign of robbery. Another pose taken into account can be of people kneeling down as shown in the figure 4 and 5.

The process identifies the image as hands up which leads to a threat being detected. For carrying out this process, the Caffe model will be used which can work with different architectures of deep learning.

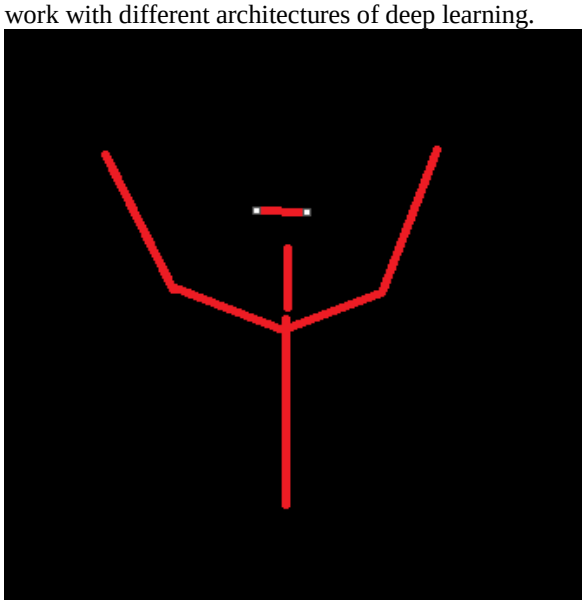

Fig 3. Hands-up pose to be detected.

Fig 4. A person sitting on his knees detected.

Fig 5. A person with his hands up detected.

D. Alert System

The requirement for an alert system arises when employees in the financial institution have to manually press the alert button and have to endanger their life. To avoid this delay and decrease the time of system alert, an autonomous system must be used which is widely used, fast, and easily accessible.

WhatsApp Messenger is widely used all over the world with more than 2 billion monthly active users globally. Python library 'pywhatkit' is a Powerful WhatsApp Automation Library capable of sending images and alerts to any user.

V. EVALUATION PARAMETERS

A. YOLO methodology

The YOLO algorithm performs image classification and image localization. In image classification, we simply get the output of whether the object in the image is a robber or not (referring to Fig 3).[1] In image localization, we get bounding boxes i.e. the position of the object along with the class. We describe the bounding box using 4 descriptors:

- Center of the bounding box
- The width
- The height
- A value c that corresponds to a class (like a gun, mask)

Another descriptor p which represents the probability of whether the object is present inside the bounding box or not. The input image is split into cells and each cell will produce bounding boxes out of which many of them will not contain the desired object that we need to detect. Therefore, we need to predict the value p. [2] The Convolution Neural Network (CNN) predicts the probabilities of various classes and bounding boxes simultaneously.

In the next step, we remove the extra bounding boxes which do not contain objects with high probability. So, in this way we create a custom dataset model using about 3000 images for detecting a robber by making three classes which consist of helmet, gun and mask. The YOLO processes images separately and it acts like a tracking system when attached to a surveillance camera, that detects objects as they move around and mold into different shapes and sizes.

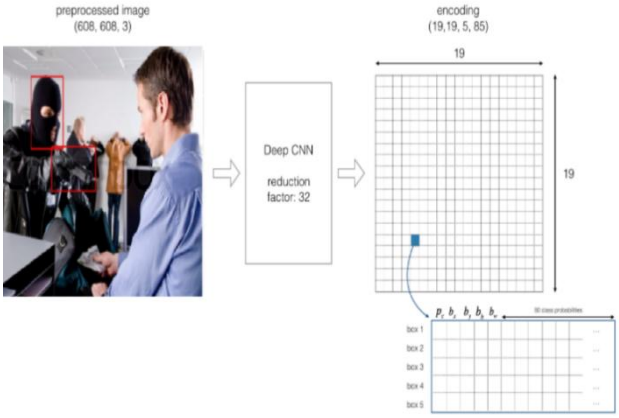

Fig 6. Image converted to cells for producing bounding boxes.

Fig 7. The process of non-max suppression is performed.

Fig 8. Robber wearing mask and holding guns.

B. Pose detection methodology

Pose detection uses the angle between joints to calculate the position values. The angle is given as:

$$\text{Angle} = |\arctan m| \quad (1)$$

Where, arc tan = trigonometric inverse tan function and m = slope. The slope m is given by:

$$m = (y2 - y1) / (x2 - x1) \quad (2)$$

Here, x1, x2, y1, and y2 represent coordinate pairs of two points. The angle values calculated are saved as arrays [13].

This process classifies the body pose into two classes:

- Hands up
- Knelt down

Class is identified by the average angle difference between all joints.

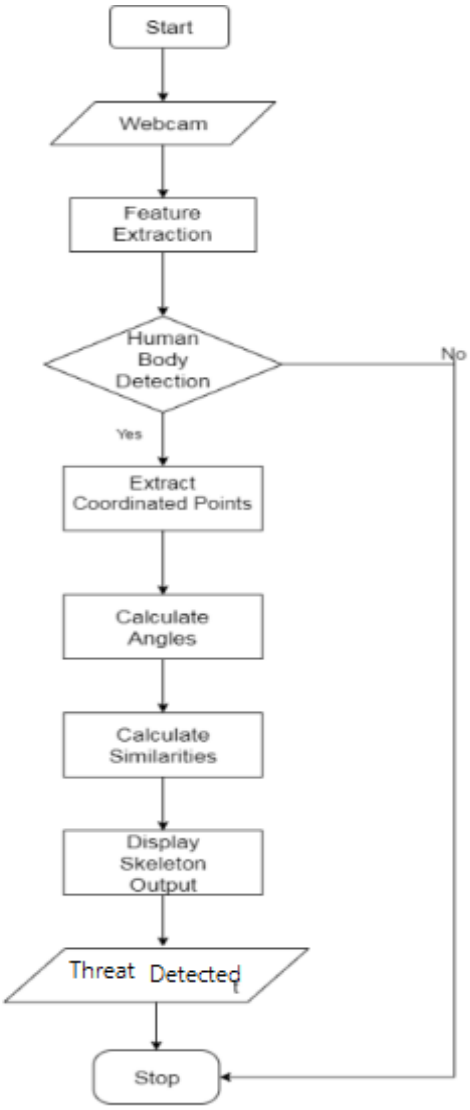

Fig 9. System flow for body pose detection to understand the potential threat.

C. Alert system trigger

The alarm notification to the law enforcement authorities will be sent based on the output of the following equation:

$$s = h / num * thresh + k / num * thresh + a[0] * thresh // 2 + a[1] * thresh // 2 \quad (3)$$

Here, a [0] = mask or helmet, h = hands up, k = kneel down, a [1] = gun and num = Total number of individuals detected in a particular frame

When s > = thresh, then criminals will be detected and the WhatsApp will send the alert to the police.

VI. IMPLEMENTATION AND RESULTS

Fig 10. Frame detected from the live feed and marked by OpenCv

Fig 11. Robber detected with the help of YOLO trained model

Fig 12. Hands-up pose detected

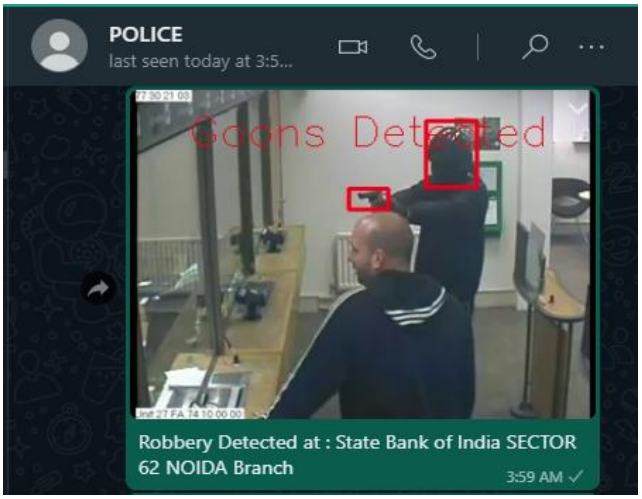

Fig 13. Robbery alert is sent to police as soon as the threshold crosses.

Fig 14. Potential Robbery alert is sent to the police with the location address.

VII. CONCLUSION

We will be able to detect and prevent theft, burglary, and robbery in financial institutions without endangering human life. The managing staff won't have to perform any action under tension and stress in the bank environment.

The system would autonomously detect criminal activity using YOLO and Caffe model. Therefore, the authorities will be informed of the bank's location without any alarm sounds.

So, if any of the two items are detected out of 3 object classes (gun, mask, and helmet) or the hands up and knees down positions are identified by Caffe model that would result in alerting the authorities.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[2] M. S. Munagekar, "Smart Surveillance system for theft detection using image processing", International Research Journal of Engineering and Technology. Aug.-2018

[3] Michał Grega, Andrzej MatiolańskiOrcID, Piotr Guzik and Mikołaj Leszczuk "Automated Detection of Firearms and Knives in a CCTV Image" https://www.mdpi.com/1424-8220/16/1/47

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015

[5] D. M. Dinama, Q. A'yun, A. D. Syahroni, I. Adji Sulistijono and A. Risnumawan, "Human Detection and Tracking on Surveillance Video Footage Using Convolutional Neural Networks," 2019 International Electronics Symposium (IES), 2019, pp. 534-538, doi: 10.1109/ELECSYM.2019.8901603.

[6] "Comparison of YOLO v3, Faster R-CNN, and SSD for Real-Time Pill Identification" Lu Tan1, Tianran Huangfu1, Liyao Wu1, Wenying Chen1* https://www.researchsquare.com/article/rs-668895/latest.pdf

[7] M. Naveenkumar and V. Ayyasamy, "OpenCV for Computer Vision Applications", Proceedings of National Conference on Big Data and Cloud Computing (NCBDC'15), pp. 52-56, March 2015.

[8] Culjak, D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, "A brief introduction to OpenCV," 2012 Proceedings of the 35th International Convention MIPRO, 2012, pp. 1725-1730.

[9] M. A. Vinith, G. Pradeep and B. Priya, "An Approach for Detecting and Identifying Suspected Weapons Using YOLO Algorithm," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 478-480, doi: 10.1109/ICSPC51351.2021.9451686.

[10] M. Putra, Z. Yussof, K. Lim, and S. Salim, "Convolutional neural network for person and car detection using yolo framework," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 1-7, pp. 67–71, 2018.

[11] Vaishali, S. S. (2019). Real-time object detection system using CAFFE model. International Research Journal of Engineering and Technology (IRJET) Volume, 6.

[12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of ACM Multimedia, pages 675–678, 2014.

[13] M. C. Thar, K. Z. N. Winn and N. Funabiki, "A Proposal of Yoga Pose Assessment Method Using Pose Detection for Self-Learning," 2019 International Conference on Advanced Information Technologies (ICAIT), 2019, pp. 137-142, doi: 10.1109/AITC.2019.8920892.