Acta Scientific COMPUTER SCIENCES

Volume 5 Issue 5 May 2023

Research Article

Real-Time Threat Detection Using the Yolo Version-4 Algorithm

Subhani Shaik, V Kakulapati*, Saadiq, Ontela Sanjay and Krishna Reddy

Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad, Telangana, India

***Corresponding Author:** V Kakulapati, Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad, Telangana, India.

Received: March 25, 2023

Published: April 12, 2023

Abstract

Big data applications are consuming the bulk of the space in industry and research. CCTV camera video streams are just as important as social media data, sensor data, agricultural data, medical data, and data produced from space research when it comes to big data. Surveillance videos provide a substantial contribution to unstructured big data. All places where security is a concern have CCTV cameras installed. It indicates that manual surveillance is inconvenient and time-consuming. Depending on the scenario, security may be defined in a variety of ways, such as detecting theft, detecting violence, estimating the risk of an explosion, and so on. The phrase "security" in crowded public places applies to almost any uncommon incident. In addition to assessing whether the captured movements are odd or suspicious, it demands a workforce and constant attention. Much of the research in the literature review suggested implementing surveillance using hardware and software tools that take video as input and require massive datasets. As it includes group activities, detecting violence among them is tough. Due to various real-world constraints, detecting anomalous or aberrant behaviour in a crowd video scene is extremely challenging. This work begins with item recognition in a crowded area. The main goal of this application is to identify weapons in the surrounding area, such as guns, knives, and fire, and to notify management of the potential threat by sending a screenshot to the user interface. This could be a good way for security and law enforcement staff to find out about the weapon in the surveillance.

Keywords: Yolo Version-4 Algorithm; Surveillance Videos; CCTV Camera; Threat Detection; Sensors; Violence; Risk; Hardware

Introduction

Due to the increase in crime, security is always a top priority in every location. Our most basic human right, the right to life, is now in danger due to gun violence. Gun violence is a daily tragedy that has an impact on people's lives all around the globe. Every single day, these weapons are used in acts of aggression that result in at least 400 fatalities, if not more. The public's worry about security is greatly impacted by these lethal weapons. Due to the growing need for the protection of safety, security, and any personal property, video surveillance systems are needed and deployed. These systems can recognize and figure out what's going on in scenes and unusual situations, which is very important for intelligence monitoring.

At this time, the number of offensive activities is steadily expanding. With this increase, preventing and detecting them is more crucial than ever. Surveillance cameras are becoming more common in public spaces. A significant number of videos are produced and archived over time. Because it takes a huge workforce and continuous monitoring to keep track of these surveillance videos, police are finding it increasingly difficult to determine whether the instances are suspicious [14,15]. As a result, there is a growing demand for precision automation of the process. It's also necessary to identify the frame that's being used. As a result, there is a growing demand for precision automation of the process. It's also necessary to identify the frame that's being used. By comparing the problem phrasing and approaches to solving used in outlier detection studies with those used in computerized supervision, this study hopes to bring cohesion to two otherwise disparate activities [1]. Anomaly detection in automated surveillance falls under the category of behavior categorization problems, which may be simplified to a problem with two or one class. Surveillance-related data is gathered through ambient detectors.

In the automated surveillance procedure, lance targets are used to detect weapons, with some behaviors thought to be abnormal. After that, a feature extraction process is used for the raw sensor data. The generated features are given into a modeling system, which uses a learning mechanism to determine whether or not the observed behavior is typical [2]. Using multiple Deep Learning models, this research seeks to detect and classify weapons in the frame. This project divides videos into chunks. A detection alert is raised in the event of a threat, indicating that the weapon has been discovered at a certain moment [13].

To develop the most efficient forecast with a high mini-batch size, several GPUs are required; achieving this with a single GPU renders training very sluggish and unfeasible. With a smaller mini-batch size and the ability to train an object detector on a single GPU, Yolo v4 solves this problem.

The goal of object automatic detection is to determine the kind and position of a particular item in a picture. For navigation, protection signals, mobility modeling, and other choices, it is extensively utilized in a variety of autonomous AI technologies.

Deep learning methodologies are used to solve the problems at hand, resulting in astonishing results. CNN (Convolution Neural Network) and Route (Recursive Neural Network) were employed in this study. A CNN is a form of neural network that may be used to generate sophisticated feature maps from the input. The complexity of the input may be reduced by the extraction of high-level feature maps [3]. Modern object recognition models include a large number of factors and, as a result, take a considerable amount of time to properly train. Transfer learning improves this problem by looking at a previously taught model for a set of categorized inputs, such as Image Net, and then retraining it with fresh weights for various new classes [4]. As a result, it's mostly employed as a tool for predicting. The necessity to convey meaning to the collected frames in the recording drove the adoption of this neural network in this investigation. This system's output is used to do real-time surveillance of various companies' CCTV cameras in order to prevent and identify any weapon in the frame.

The remaining work is composed of main components: a brief description of the related works discussed in Sec 2. Sec 3 summarizes the framework of the proposed system. Sec 4 offers a brief overview of the proposed methodology sec 5 describes the implementation analysis of our system, Sec 6 provides an overall discussion and Sec 7 explores a synthesis of the main conclusions followed by future enhancement.

Literature Survey

Handheld weapons, such as guns, pistols, and revolvers, are used in the majority of criminal operations nowadays. Several polls have revealed that the handgun is the most commonly utilized weapon for a variety of crimes such as burglary, rape, and so on. As a result, automatic gun recognition is a critical necessity in the current environment, and this research uses Convolution Neural Networks [5,6] to offer automated gun detection from a congested scene (CNN).

At the moment, crime rates are on the rise, and law enforcement agencies are having more difficulty enforcing the law. Some tactics have been tried in the past few decades for identifying crime via video surveillance, and manually detecting crime is a difficult operation that carries some risk. A deep learning algorithm was introduced for spotting criminality in video scrutinizing aperture to overcome this issue. The Complexity Clear System was employed to increase detection accuracy in this suggested technique [7,8]. As a result, the accuracy of detecting crime has improved. Furthermore, the created error is considerable in the combination of characteristics, so it cannot give reliable detection.

System framework

The framework is the process of arranging a system's architecture, modules, interfaces, and data to achieve certain goals. It may be defined as a product improvement application of a system's idea. Object-oriented evaluation and design methodologies are speedy becoming the maximum popular ways of developing pc systems [9].

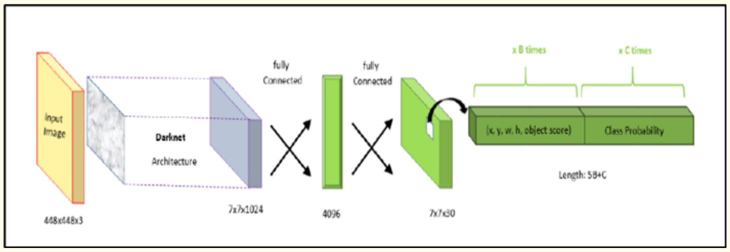


Figure 1: Darknet Architecture.

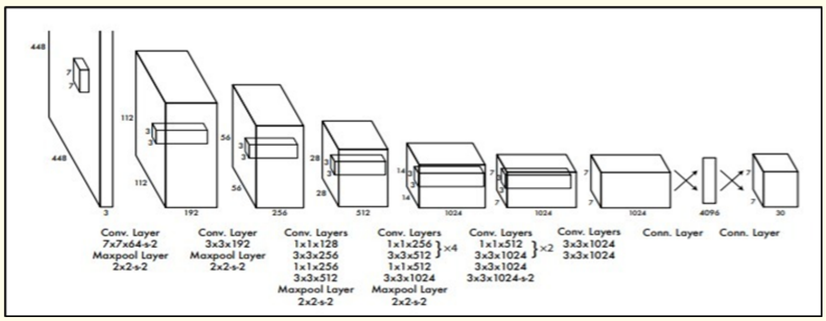


Figure 2: Darknet Architecture.

The suggested model includes a precise definition for detecting suspicious behavior. The rate of archival crime is rapidly growing. It is quite difficult for a human to keep an eye on every location on the planet to avert disasters and illegal. As a result, we tend to propose our model. InceptionV3 is a convolution neural network that has been pre-trained. For initial classification, a recurrent neural network is utilized, and for final classification, a recurrent neural network is employed suspicious activity detection It is functioning admirably in terms of higher accuracy. The intended model's architecture is depicted in the diagram below figure.

Methodology

The objective of this study was to develop a unified security framework that would use IP cameras to monitor suspicious activity and alert security personnel to the scene with instructions on how to handle the issue if identification could be established. We propose a model that trains a computer to identify firearms and alerts a human administrator if it detects a gun or other potentially lethal weapon. Additionally, we have a mechanism that automatically locks doors if the shooter seems to have a deadly weapon. With the use of IP cameras, we may also provide a live feed to security professionals so that they may take immediate action.

We've also built an information system to track all of the drills and transmit impact operations in the urban areas in case of a future crisis. One such result is the establishment of a repository for documenting all of the happenings for use in case of future emergencies. The three components of our research's overarching approach are shown in figure 1.

The most crucial part of every application is having wanted and acceptable data to train machine learning models on. So, we culled several images from Google by hand. Some representative images are shown in figure 3. For each kind of weapon, we collected at least 50 images. Google-images-download is a great tool for obtaining high-quality photos for use in one's own data collection efforts. These pictures were placed in a folder labeled "images." If the needed photographs have different filename extensions, it will be more challenging to train with them, which will increase the likelihood of errors being made.

Object detection is largely concerned with computer vision, which entails recognizing things in electronic pictures [10,11]. Object identification is one of the fields that has profited greatly

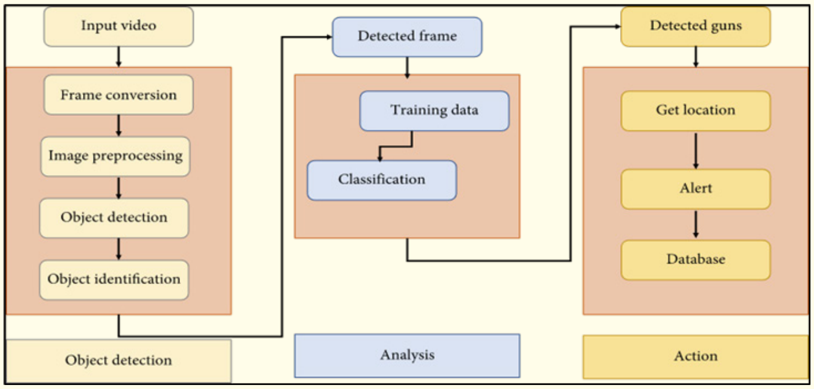


Figure 3: Framework of the proposed method.

from recent advances in deep learning. YOLO is a kind of object detector that has been pre-trained. It's a CNN simulation. A CNN is a deep learning method that can take a raw input image and assign learnable weights and biases to the image's many aspects/objects. A convolution layer in a CNN model is responsible for withdrawing high-level characteristics from the input picture, such as edges [12,13]. This works by repeatedly applying the kernel filter on the raw picture. Activated maps or feature maps can be said to be the outputs of these.

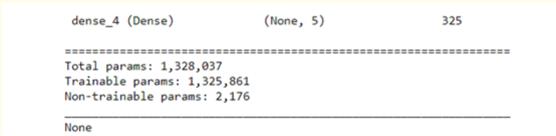
The model also detects the emotion of the gun holder to detect the threat level to the people among them. If the emotion of the gun holder is happy, afraid, or angry then the threat level is maximum, to calculate the emotion of the gun holder we calculate the distance from the weapon origin position to the face origin position if it is less than 150 than the emotion belongs to gun holder else the face is passed.

The Emotion detection model consists of 8 Convolution layers with a Batch Normalization layer after each Convolution layer and a Dropout layer after every two Convolution layers. The convolution layer is containing a set of filters that are used for training, size of the input image decreases after it passes through each Convolution layer using the following formula.

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1$$

n_{in}: number of input features
n_{out}: number of output features
k: convolution kernel size
p: convolution padding size
s: convolution stride size

Layer (type)	Output Shape	Param #
conv2d_8 (Conv2D)	(None, 48, 48, 32)	320
batch_normalization_10 (BatchNormalization)	(None, 48, 48, 32)	128
conv2d_9 (Conv2D)	(None, 48, 48, 32)	9248
batch_normalization_11 (BatchNormalization)	(None, 48, 48, 32)	128
max_pooling2d_4 (MaxPooling2D)	(None, 24, 24, 32)	0
dropout_6 (Dropout)	(None, 24, 24, 32)	0
conv2d_10 (Conv2D)	(None, 24, 24, 64)	18496
batch_normalization_12 (BatchNormalization)	(None, 24, 24, 64)	256
conv2d_11 (Conv2D)	(None, 24, 24, 64)	36928
batch_normalization_13 (BatchNormalization)	(None, 24, 24, 64)	256
max_pooling2d_5 (MaxPooling2D)	(None, 12, 12, 64)	0
dropout_7 (Dropout)	(None, 12, 12, 64)	0
conv2d_12 (Conv2D)	(None, 12, 12, 128)	73856
batch_normalization_14 (BatchNormalization)	(None, 12, 12, 128)	512
conv2d_13 (Conv2D)	(None, 12, 12, 128)	147584
batch_normalization_15 (BatchNormalization)	(None, 12, 12, 128)	512
max_pooling2d_6 (MaxPooling2D)	(None, 6, 6, 128)	0
dropout_8 (Dropout)	(None, 6, 6, 128)	0
conv2d_14 (Conv2D)	(None, 6, 6, 256)	295168
batch_normalization_16 (BatchNormalization)	(None, 6, 6, 256)	1024
conv2d_15 (Conv2D)	(None, 6, 6, 256)	590080
batch_normalization_17 (BatchNormalization)	(None, 6, 6, 256)	1024
max_pooling2d_7 (MaxPooling2D)	(None, 3, 3, 256)	0
dropout_9 (Dropout)	(None, 3, 3, 256)	0
flatten_1 (Flatten)	(None, 2304)	0
dense_2 (Dense)	(None, 64)	147520
batch_normalization_18 (BatchNormalization)	(None, 64)	256
dropout_10 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 64)	4160
batch_normalization_19 (BatchNormalization)	(None, 64)	256
dropout_11 (Dropout)	(None, 64)	0



```
 dense_4 (Dense)             (None, 5)                 325

=================================================================
Total params: 1,328,037
Trainable params: 1,325,861
Non-trainable params: 2,176
_________________________________________________________________
None
```

Figure a

The batch Normalization layer normalizes the contributions to a layer for every mini-batch. The dropout layer is used to remove connections between layers so that the model doesn't overtrain. The dense layer is a simple layer in which the layer receives information from the previous layer.

Results and Analysis

Object detection is largely concerned with computer vision, which entails recognizing things in electronic pictures. Object identification is one of the fields that has profited greatly from recent advances in deep learning. All places where security is a concern have CCTV cameras installed. It indicates that manual surveillance is inconvenient and time-consuming. Depending on the scenario, security may be defined in a variety of ways, such as detecting theft, detecting violence, estimating the risk of an explosion, and so on.

Figure 4: Arms Detection in CC TV Surveillance.

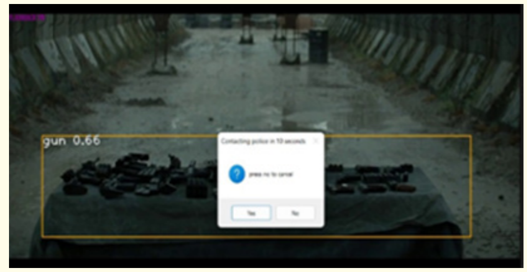


Figure 5: Multiple Arms detection.

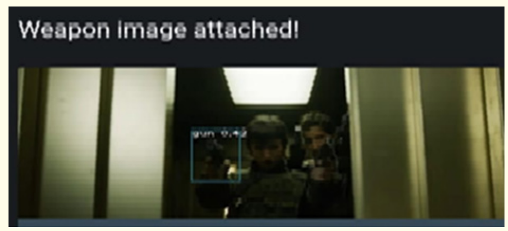


Figure 6: An Email sent to the Authorities.

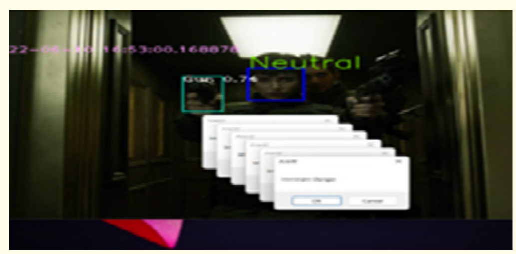


Figure 7: Emotion detection of the gun holder.

Figure 5 discusses CCTV surveillance for arms detection in a single point, figure 6 gives information about multiple arms detection, and finally figure 6 for mobile CCTV captures the image, it can observe through the mobile application.

Conclusion

We establish a paradigm that enables machines or robots to see firearms and alerts humans when they see a gun at the edge of the screen. Current surveillance capabilities must be enhanced as soon as possible with more resources to put human operators to the test. Smart surveillance systems will replace the current infrastructure as low-cost storage, video infrastructure, and higher video processing capability become generally accessible. Digital monitoring systems in the appearance of robots will ultimately completely replace existing surveillance systems as more and more affordable computers, video infrastructure, high-end technologies, and improved video processing become widely accessible.

Future Work

We may infer from this experiment that by using Deep Learning algorithms, we can identify suspicious actions taking place around us. The approaches we came across are described in depth and extensively examined to determine their benefits and drawbacks. Because not all types of actions are recognized, the suggested system offers a lot of room for improvement in the future. As a result, it may be developed such that it can identify a wide range of behaviors from live CCTV footage.

Bibliography

1. Radhika C., *et al*. “Face Recognition Based Attendance System Using Machine Learning Algorithms”. Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART (2018).

2. Omar Abdul., *et al*. “Class Attendance Management System Using Face Recognition”. 2018 7th International Conference on Computer and Communication Engineering (ICCCE) IEEE (2018).

3. Adrian Rhesa Septian Siswanto., *et al*. “Implementation of Face Recognition Algorithm for Biometrics Based time Attendance System “Center for Information Communication Technology Agency for the Assessment Application of Technology (PTIKB-PPT) Teknologi 3 BId., 3F, PUSPIPTEK Serpong, Tangerang, INDONESIA, 15314.5ITM Web of Conferences 32,02001 (2020).

4. Jinsu Kim., *et al*. “Face Recognition Enhancement by Employing Facial Component Classification and Reducing the Candidate Gallery Set. Department of Computer Engineering, Sejong University, Seoul, 143-747, Korea (sbmoon@sejong.ac.kr).

5. Nusrat Mubin Ara., *et al*. “Convolutional Neural Network (CNN) Approach for Vision Based Student Recognition System”. 2017 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December (2017).

6. Xin Xiong Liu and Wanru Wang. “User-Centered Design of Attendance Record System Based on Mobile Terminals” 2017 International Symposium on Computer Science and Intelligent Attendance System in College Using Face Recognition and NFC”. *International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology* 6.6 (2017): 14-21.

7. Rekha E and Ram Prasad. P. “An Efficient Automated Attendance Management System based on Eigen Face Recognition”. Department of Electrical Engineering, Amity University, Dubai-UAE.

8. Shu Zhan Toru Kuriharak and Shigeru Ando “Facial Authentication System Based on Real-Time 3D Facial Imaging By Using Correlation Image Sensor” School of Computer and Information Department of Information Physics and Computing Hefei University of Technology The University of Tokyo Hefei, Anhui, 230009, China 7-3-1 Bunkyo-Ku, Tokyo113-8656.

9. Asri Nuhi., *et al*. “Smart Attendance System using QR code”. 9th Mediterranean conference Embedded Computing, Budva, Montenegro (2020).

10. MA Meor., *et al*. “Centre for Telecommunication Research and Innovation FakultiKej. ElektronikdanKej. Komputer Universiti Teknikal Malaysia Melaka Hang TuahJaya, Durian Tunggal 76100, Melaka, Malaysia (2014).

11. Amena Khatun., *et al*. “Design and Implementation of Iris Recognition Based Attendance Management System”. ICEEICT Jahangirnagar University, Bangladesh (2015).

12. Shreyak Sawhney., *et al*. “Real Time Smart Attendance System using Face Recognition Techniques” in Amity University Uttar Pradesh, Noida, (2019).

13. Awais Ahmed. “LBPH-based Improved face recognition at low Resolution”. UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA 2018 IEEE.

14. “Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods” Teddy Montoro, Media A. Ayu, Suhendi Sampoerna University, Jakarta, Indonesia (2018).

15. HAO YANG and XIAO FENG HAN. “Face Recognition Attendance System Based on Real-Time Video Processing” is supported in part by the Basic Public Welfare Research Project of Zhejiang Province under Grant LGF 20H 180001, (2020).