IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12214

Militant Intrusion and Weapon Detection with Voice assistance using Machine Learning

Chetana Srinivas[1],Deepak H V[2],Abhishek M V[3],Pavan G R[4],Alur Malkari Sumanth[5]

Associate Professor, Department of Computer Science and Engineering, EWIT Bangalore, India[1]

Research Intellect, Department of Computer Science and Engineering, EWIT Bangalore, India[2-5]

Abstract: In opposition to crime and criminals. The police are becoming less willing to respond to crime scenes unless there is visible confirmation, either by manned patrols or by electronic images from the surveillance cameras. The current systems do not classify routine and abnormal events.The proposed work is used for a variety of reasons, including live tracking, monitoring, classifying weaponry, and surveillance. In this work, real time image processing techniques are used to extract live surveillance footage from monitoring and identifying unusual events.The proposed project contains three processing modules. The first processing module uses Convolutional Neural Networks (CNN) for object identification, the second processing module handles the classification of weapons, and the third processing module handles monitoring and alert functions. A circular area will be monitored by CCTV, which will operate and be managed automatically. Before being implemented in such an environment, shape detection algorithms and object detection algorithms have been tested for accuracy in detection and analysis of processing time. The results provide the best accuracy in matching weapon and object types with names and shapes in predefined databases like ALEXNET. The proposed work will significantly lower crime rates, increase security in some regions, and shorten the time it takes to apprehend offenders.

Keywords: Convolution Neural Network (CNN), Video Surveillance, Voice Assistance, Weapon Detection, Faster Region based Convolution Neural Network (RCNN),picture segmentation.

I. INTRODUCTION

Closed circuit television (CCTV) systems are being installed in a growing number of offices, housing developments, and public locations. In India, there are currently one million CCTV cameras in use. Human factors limit the number of camera views that one operator can keep track of which results in a tremendous workload for the CCTV operators.Monitoring gets more difficult when there are more CCTV cameras present since it is tougher to regulate,observing,detecting,recognizing, and identifying the individuals and circumstances that could be dangerous to other people and property. Automated image-understanding algorithms would not take the position of overworked human operators; rather, they would warn them when a potentially dangerous scenario arises. When someone is openly carrying a weapon (such as a knife or a gun), it is a clear sign that things might go deadly. Even if some nations permit open carry of guns, Even so, it's crucial to get the CCTV operators' attention in this case so they can assess the existing circumstance. In recent years, there have been more instances when dangerous drugs have been used. Automated video surveillance methods have initialized to appear in now days, especially for application in intelligent transportation systems (ITS). We have concentrated on the unique problem of automated recognition and detecting of hazardous situations suitable generally to any CCTV device. Our goal is to automatically detect intruders and dangerous objects, such as fire arms and knives, which are the very often used and deadly weapons. Examples of warning signs to which the human operator must be aware include the look of such things being held in the hand. If a potentially harmful scenario arises, notify the human operator. When someone is openly carrying a weapon (such as a knife or a gun), it is a clear sign that things could go dangerous. Even though some nations permit open carry of firearms, In this circumstances, it is yet advisable to get the CCTV operators attention so they can assess the present circumstance. In recent years, there had been an upsurge in the numerous cases involving the use of automated video surveillance techniques have begun to develop, primarily for use in intelligent transportation systems (ITS). We have concentrated on the particular task of automated risk detection and recognition, which is suitable to any CCTV system in general. Automated detection of intruders and dangerous weapons, as well as fire arms and knives, the very widely used and lethal weapons, is the issue we are trying to solve. Control, identify, and discover individuals and circumstances that might endanger other people and property are examples of warning indicators that the human operator needs to be made aware of, but it becomes more difficult to monitor when there are many CCTV cameras. Employing automatic image-understanding algorithms as a solution to the issue of overburdening the human operators will notice them in case of a potentially harmful state n is present rather than taking the position of the human operator. When someone is openly carrying a weapon (such as a knife or a gun), it is a clear

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12214

sign that things might go deadly.Even if some nations permit open carrying of guns, it is yet necessary to get the CCTV operators' focus in this circumstance so they can examine the current situation. In present days, there had been an upsurge in the numerous cases involving the usage of In recent years, automated video surveillance techniques have begun to develop, primarily for use in intelligent transportation systems (ITS). We have concentrated on particular challenge of automated risk detection and recognition, which is applicable to any CCTV system in general. Automated detection of intruders and hazardous weapons, including fire arms and knives, the very widely using and lethal weapons are the issue we are trying to solve. An example of a warning indication that the human operator has to be made aware of is the look of such things being held in the hand.

II. IMPLEMENTATION

The modules incorporated in this project are:
1. Dataset upload
2.Pre-processing data
3.Extracting dataset
4.Splitting dataset into training and testing
5.image processing
6.Video converting into number of frames
7.Applying models
8.Voice assistance

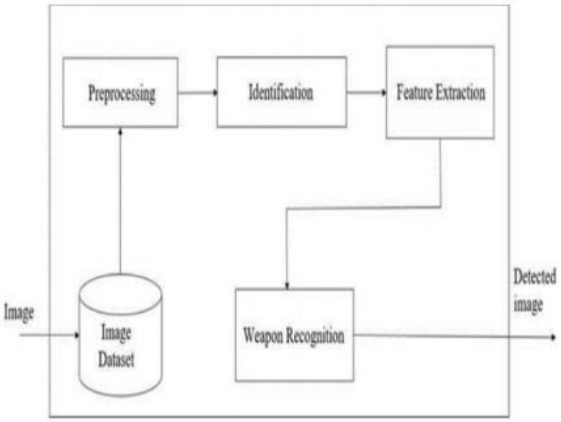


Figure1:System architecture of Weapon detection

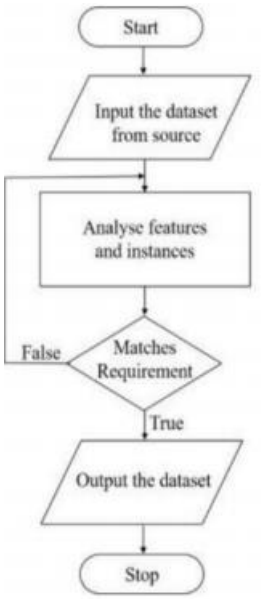


Figure 2: Flow chart for the module [6]

The flow chart for grouping information is as portrayed within the figure 2 the information set is collected from a supply and a whole analysis is dispensed. The image is chosen to be used for training/testing functions provided that it matches our necessities and isn't recurrent.

2.1 TECHNOLOGIES USED

2.1.1 SSD Algorithm
A technique for identifying police-related objects in images using one deep neural network at a time. For each feature map point, the output house of bounding boxes from our approach, known as SSD, is discretized into a the AIML intelligence environment is updated to include the data from gTTS.

2.2.2 Faster R_CNN
A single-stage model that is trained from beginning to end could be a faster R-CNN. When compared to antiquated methods like Selective Search, it uses an entirely new region proposal network (RPN) to provide regional proposal ideas. Utilizing the ROI Pooling layer, it takes each area proposal and extracts a fixed-length feature vector.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12214

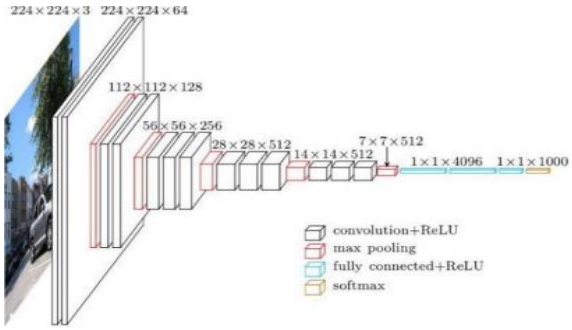


Figure 3: Layers in CNN Architecture[4]

Figures 3 shows the layers of the faster RCNN architecture. It has 2 networks: a network for object identification and an RPN network for producing region proposals. It employing the selective finding methodology to make region of proposals. By RPN network, anchors or an region boxes are ranked.[6]

2.1.3 YOLO Algorithm

YOLO is a formula that provides period object detection using neural networks. The YOLO formula is an Associate in Nursing formula based regression that predicts categories and bounding boxes for the entire image in one run rather than just the visually appealing portion.[5]

2.1.4 Weapon detection

The discovery of irregular, unexpected, unpredictable, uncommon events or things that are not deemed to be regularly occurring events or regular items in a pattern or items included in a dataset and are thus dissimilar from current patterns is known as weapon or anamoly detection. A pattern that deviates from a set of expected patterns is called an anomaly. Anomalies therefore rely on the phenomenon under study. In order to identify occurrences of different categories of items, object detection employs feature extraction and learning techniques or models. Implementation ideas emphasise precise gun detection and classification accuracy is also a worry because a false warning could trigger unfavourable reactions. Making the optimal trade-off between accuracy and speed required selecting the appropriate strategy. The process for detecting weapons using deep learning is shown in Figure 1. From the input video, frames are extracted.[4]

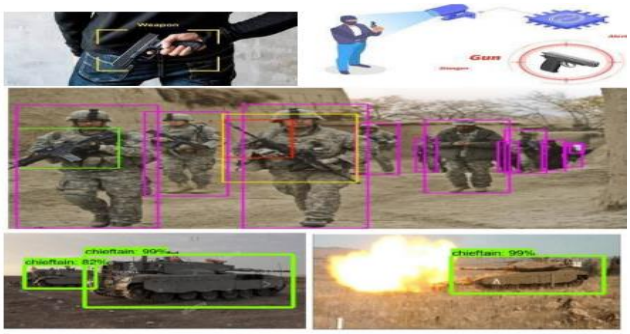


Figure 4: Militants and multiple weapon detection

2.1.5 Video Surveillance

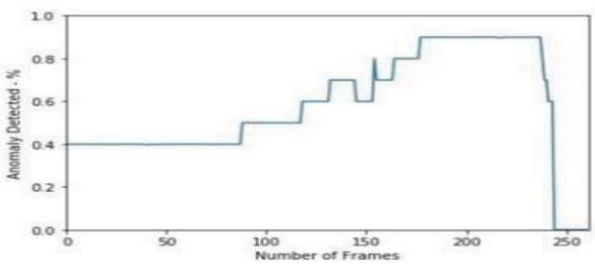


Figure 5: Finding anomalous events in video frames [2]

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12214

To check if somebody is "jumping" the fence, a specific anomaly is set to finding any anomalies inside and around the fence. This is accomplished by applying the mathematical model to extract the masked binary(0 or 1) value of the person identified in the video frame by frame. To determine the person's original and current positions, the summed location values from each frame will be recorded. This will make it easier to tell if someone is trying to jump over the fence or is already jumping over it when it is in its normal position. Figure 6 depicts a human. going over the fence, figure 7 depicts the identification of the individual using probability values, and figure 5 displays the anomaly value plotted for each frame of the video from the beginning to the finish. According to figure 5's percentage of anomalies discovered for individual frames, a person-climbing gate anomaly event happens between frames 150 and 225. Additionally, we can observe that the anomaly value is a constant 0.4% from frames 0 to 88. This is because the individual is on a set of stairs,which is a somewhat higher
location than usual.[2]

Figure 6: Detecting a human in an input video frame [2]

Figure 7:Detecting a human in a video frame output[2]

2.1.6 Image Segmentation

Algorithms for image segmentation that use edge, shape, and point detection. The two primary picture segmentation techniques are the active appearance models and the Harris interest point detector. These algorithms function well when the photos that contrast. Any image noise makes it difficult to identify the image's corner and form..[3] Convolutional neural networks are the primary foundation for deep learning object detection methods (CNNs). For training and testing purposes, neural networks need access to a variety of image datasets and information about an object's position. The two primary neural network algorithms for detecting knives and pistols are faster R- CNN and YOLO. Faster R-CNN has been shown to be more accurate, 93%, although YOLO achieved a superior performance of around 30+ frames per second, according to the work(s). [3]

III.METHODOLOGY

3.1 Database
We constructed and tested our method using the benchmark weapons database,[1] the Internet Movie Firearms Database (IMFDB).

Internet movie Fire arms Database (IMFDb)

An IMFDb is DB of firearms that have been seen or used in movies, tv programmes, video games, as well as animation. The pictures of weaponry were taken from Japanese cartoons, TV shows, video games, and Hollywood films. Assault weapon, battle rifle, bullpup, carbine, flare gun, grenade, mine, missile launcher, mortar, shotgun, sniper rifle, submachine gun, handgun, revolver, and underwater weapon, mortar, etc. are included in the gun category. Although there are many other classifications of firearms in IMFDb, we have only included revolvers, rifles, and shotguns. Figure1 displays a selection of the database's favourable examples. The negative images are randomly selected from a variety of genres, including animals, flowers, and landscapes, on the internet.[1]

3.2 Deep Learning Model

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

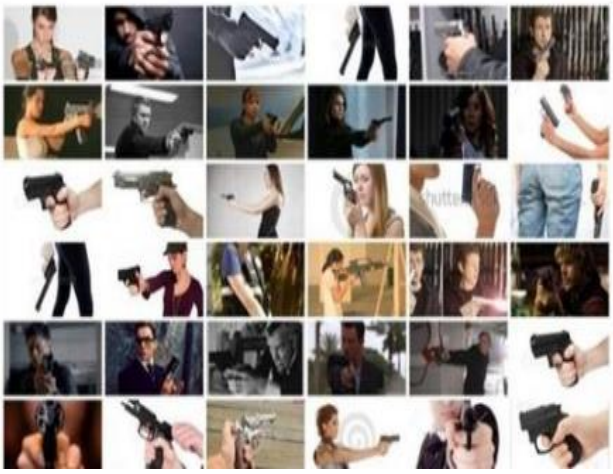
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12214

Modern Convolutional Neural Networks (CNNs) may be produced without the need of a graphics processor unit(GPU),we used MatConvNet, a MATLAB toolkit. In this investigation, we employed a pre-trained classification model based on the Image Net dataset called VGG-16 (nearly 1.28 million pictures across 1,000 collective object classes). There are two versions of VGG Net: VGG-16 and VGG-19. There are 16 convolutional layers in the VGG-16 architecture, with millions of parameters for each. 3 completly connected layers plus an linear layer with the activation function "Softmax" make up the output of the model. Before RELU activation is incorporated to all convolutional layers of the VGG-16, dropout regularisation is employed in a fully connected layer. Prediction loss reduction is widely used for training deep CNN, such as the VGG-16. Equation 1's description of the intension of training is to iteratively reducing the average loss using the input pictures x and y and matching output class labels.[1]

$$J(w) = \frac{1}{N}\sum_{i=1}^{N} Ł\left(f\left(w, x_i\right), y_i\right) + \lambda R(w) \qquad (1)$$

3.2.1 Detection of Handheld Gun by Faster R-CNN Deep Learning



8(i): Positive

8(ii): Negative

Figure 8 : (i) Sample positive photos; (ii) Sample Negative photos from the IMFDB database.[1]

N is total number of data instances (mini-batch) in individual iteration, L is the loss function, and R is the weight decay with the Lagrange multiplier, where the projected output of the network is f based on the ongoing weight w. We update the weights using the Stochastic Gradient Descent (SGD), a technique frequently employed in deep CNNs, and provide equation 2 as a result.

$$Wt+1 = \mu * w * t - \alpha * \Delta J * (wt) \qquad (2)$$

where the learning rate is r and The current weights' momentum weight is wt. If the network is trained from scratch, the weights of the network are started without any prearrangement. If the deep model is being fine-tuned, the network weights are first set to pre-learned network weights. In this study, we employed the VGG-16 that can be tuned and the same architecture's weights were used to initialise it. Picture total Database(Db) pre-training for VGG-16. MatConvNet, an deep learning tool for MatLab, was used to acquire the pre-trained VGG-16 model for this investigation.[1]

IV. VOICE ASSISTANCE

With the development of AI and intelligent assistants like Amazon's Alexa, Microsoft's Cortana and Apple's Siri the Voice recognition has becoming most commonly used. Voice recognition technology offers hands-free requests, reminders, and other fundamental operations so that users may interact with technology directly by speaking to it. The basic positions of the objects in the subject's or camera's field of view can then be provided by translating the annotated text into vocal answers. Modern deep learning models can recognise speech and identify things in photos, but they do it individually and with different models. You would probably assert that this is what AGI is supposed to be if I told you that there is a way to create a single model that can integrate these two functions, even if we are far from that.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023
DOI: 10.17148/IJARCCE.2023.12214

4.1 GTTS: SPEECH SEGMENTATION

Total speech synthesis is created via the system's gTTS engine. Text-to-Voice module of Google recognises and interprets the speech that is delivered to it using deep learning neural networks. Through the integrated mixing feature of gTTS, the speech is translated for the engine. Every data is processed by the JARVIS using the Microphone as its root, and the output is then either converted to ".mp3" or ".txt" format.[7]

4.1.1 Relocating to AIML

By parsing the varying forms to Python and subsequently into gTTS, Remapped into the AIML intelligence environment is the data from gTTS..[7] The data flow is as follows:
(a) Python Input from user: The user's spoken words are regarded as a ".mp3" input file for the Python script.
(b) gTTS from Python Interpreter: The script's ".mp3" file is further processed and sent to the gTTS engine for conversion from voice to text.
(c) AIML from gtts*:* The convocation of the user's spoken paraphrases results in the new text produced by the gTTS. The AIML module is parsed from this and attached to gTTS using Python scripting.
(d) gTTS from AIML: The AIML module produces a text-based response in response to the user's input inquiry. The resulting text is then transferred to the gTTS engine for reconfiguration, which is now performing the opposite operation for converting the text into voice.
(e) gTTS from Phython : At last, the gTTS-generated ".mp3" format is processed into a Python file and imported into the Bootstrap Kernel. The script's various APIs are then utilised to play this speech by Python, completing the speech segmentation.

V. CONCLUSION

Since the identification and categorization of weapons and militant detection are accomplished successfully, the project's results have been good. As the information contained in weapon and militant photographs, the pre-processing procedures employed to minimise the amount of data input during the classification have shown to be effective. The creation of these subsystems has achieved the goal of utilising only the important data, namely the pixels surrounding potential militant and weapon-producing areas. The procedure of data augmentation has been effective and efficient. The amount of input data required to train the CNN was insufficient. The rotation and translation of the photos allowed for the collection of a huge number of weapon examples and extremists to teach the CNN what they have learned. the picture dataset that was manually annotated using image detection with YOLOv5. Speech recognition is employed in this research to interrogate the trained model using the Loaded image. expecting the audio output to respond to the trained model's loaded image.

REFERENCES

[1] Anamika Dhillon ,Gyanendra K. Verma, Anamika Dhillon. "A Handheld Gun Detection using Faster R-CNN Deep Learning" , Proceedings of the Seventh International Conference on Computer and Communication Technology - ICCCT-2017.
[2] Vidyashree Dabbagol,Ruben J Franklin, Mohana. "Anomaly Detection in Videos for Video Surveillance Applications using Neural Networks" , 2020 4th International Conference on Inventive Systems and Control (ICISC), 2020.
[3] Arif Warsi, Munaisyah Abdullah, Muhammad Yahya, Mohd Nizam Husen. "Automatic Handgun and Knife Detection Algorithms: A Review" , 2020 fourteenth International Conference on Ubiquitous Information Management and Communication (IMCOM), 2020.
[4] Harsh Jain, Ayush Jain ,Aditya Vikram, Mohana, Ankit Kashya. "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications" 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020.
[5] F.I. Abd-AL Sahib, R F. Ghani, R F. Ghani. "Detection of the autonomous car robot using Yolo" , Journal of Physics: Conference Series, 2021.
[6] P.Shanmugapriya, Jammalamadaka mahendar kumar, Gurryugandhar reddy. "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications" 2022 International Research Journal of Engineering and Technology (IRJET) ISO 9001:2008 Certified Journal.
[7] Ravivanshikumar Sangpal, Tanvee Gawand, Sahil Vaykar, and Neha Madhavi. "JARVIS: An interpretation of AIML with integration of gTTS and Python" 2019 Second International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT).

 This work is licensed under a Creative Commons Attribution 4.0 International License