Cogni-Watch Intelligence Video Surveillance System

A.Jayasimha[1], P.Mallesh[2], N.Harsha Vardhan[3], Mohammed Afzal[4]
[1,2,3] *B.Tech Student, Department of Artificial Intelligence and Machine Learning, Sphoorthy engineering college*
[4]*Assistant Professor, Department of Artificial Intelligence and Machine Learning, Sphoorthy engineering college*

Abstract- In both industry and research, big data applications are taking up much of the available space. Video streams from CCTV cameras are one of the most common examples of big data, and they play a significant role alongside data from other sources such as social media, sensors, agriculture, medicine, and space exploration. The contribution of surveillance videos to unstructured big data is significant. CCTV cameras are installed everywhere that security is a top priority. Manual monitoring appears laborious and time-consuming. Different definitions of security can be used to different situations, such as identifying theft, detecting violence, predicting explosions, etc. The word "security" refers to nearly every kind of unusual occurrence in crowded public areas. Among these, as it entails group work, violence detection is challenging to manage Due to a number of practical limitations, analysing anomalous or aberrant activity in a crowd video scene is quite challenging. In-depth surveying is included in the study, starting with object recognition, followed by activity recognition, crowd analysis, and violence detection in a crowd setting. Deep learning techniques constitute the foundation of the majority of the publications evaluated in this study. Algorithms and models used in various deep learning techniques are compared. This survey's primary goal is to use deep learning techniques to the task of precisely counting the number of participants, the individuals involved, and the activities that occurred in a large crowd under any kind of weather. The fundamental deep learning implementation technology used in several crowd video analysis techniques is discussed in this paper. An essential topic that has not yet received enough attention in this area is real-time processing. There aren't many ways to deal with all of these problems at once. The problems discovered with the current approaches are listed and condensed. Future guidance is also provided to lessen the obstacles found.

1. INTRODUCTION

OLO suggests using an end-to-end neural network to predict bounding boxes and class probabilities simultaneously. In contrast to other object discovery algorithms that utilized classifiers for discovery, YOLO suggests employing an end-to-end neural network to predict bounding boxes and class probabilities simultaneously. By using an atypical approach to object finding, YOLO surpasses existing real-time object discovery algorithms by a significant margin and produces state-of-the-art results. Whereas algorithms such as Faster RCNN operate by employing the Region Offer Network to identify potential regions of interest and then carry out identification on those regions separately, YOLO executes all of its predictions with the assistance of a single fully Designs that employ Region Offer Networks and related subcastes ultimately result in several duplications of the same image.

One replication is all it takes to defeat YOLO. The image is divided into N grids by the YOLO algorithm, each of which has an equal dimensions section of SxS. The task of finding and localising the object it contains falls to each of these N grids. In a similar vein, these grids predict B bounding box coordinates in relation to their respective cells, as well as the item marker and likelihood that the object will be in the cell. Because cells from the image handle both discovery and recognition, this method significantly reduces computation. However, it results in many prognostications that are indistinguishable because numerous cells predict the same object with distinct bounding box predictions.

2. HOW DOES YOLO WORK?

Assume we have a picture of a dog and a cat with two bounding boxes. YOLO divides the image into a grid as its initial step. By having a grid, it is feasible to detect one object per grid cell as opposed to one object per image. It is possible to encode a vector describing each gid cell. For example, the first cell from the top-left has no object. We describe where the object class

probability is, the bounding box's width and height relative to the entire image, and the bounding box's centre coordinates, which are either 0 or 1, depending on which class represents both the dog and cat bounding boxes. A vector is made up of symbols because, in the event that the first component is equal to zero, the remaining components may have random values or not be taken into account. We will then get a vector if we take the cell that has the cat in the centre of the blue bounding box. After this process, nine vectors with dimension 7 or 3x3x7 tensor represent the entire image if we define one vector for each grid cell. This indicates that a single 3x3x7 tensor is assigned to each image sample in our data collection. We can build a training and test set and train the convolutional network using that data set, just like how YOLO is effective. Because it can forecast every item with CNN in a single forward pass, YOLO's full name is "You Only Look Once."

2.1 YOLO v7

The most recent version of YOLO, V7, offers a number of enhancements over earlier iterations. The utilisation of anchor boxes is among the primary enhancements. Anchor boxes are a collection of pre-built boxes with various aspect ratios that are used to identify variously shaped objects. Because YOLO v7 employs nine anchor boxes, it can identify a greater variety of item sizes and shapes than earlier iterations, which helps lower the amount of false positives. YOLO v7 has made significant progress with the addition of a new loss function known as "focal loss." The standard cross-entropy loss function, which is known to be less successful at identifying small objects, was utilised in earlier iterations of YOLO. In order to combat this problem, focal loss downweights the loss for properly categorised instances and emphasising the challenging cases—the difficult-to-detect objects. Additionally, YOLO v7 boasts a greater resolution than earlier iterations. YOLO v3 uses a resolution of 416 by 416 pixels; this is lower than the 608 by 608 pixels that it processes photos at. YOLO v7 can detect tiny objects and has a higher overall accuracy thanks to its higher resolution. It should be mentioned that two-stage detectors like Faster R-CNN and Mask R-CNN, which often achieve greater average precision on the COCO dataset but also require longer inference times, are more accurate than YOLO v7.

2.1.1 YOLO v8

The publication of YOLOv8, a model that specifies a new standard for computer vision, improves the field. state of the art in terms of instance segmentation and object detection. In addition to enhancements to the model architecture, YOLOv8 provides developers with a more user-friendly interface for utilising the YOLO model through a PIP package. The most recent and advanced YOLO model, YOLOv8, is applicable to tasks like instance segmentation, object detection, and image classification. The company Ultralytics, who also developed the well-known and industry-defining YOLOv5 model, is the creator of YOLOv8. Many architectural and developer experience enhancements and modifications over YOLOv5 are included in YOLOv8. As of the time of writing this post, Ultralytics is actively working on new features and responding to community input for YOLOv8. When Ultralytics does, in fact, release As a model, it has sustained support since the organisation collaborates with the community to enhance the concept.

YOLOv8 Tasks:-

Image Classification

1. Classification of Images Classification entails classifying an image as a whole without focusing on any particular object that may be present.
2. High-Value Video Capture: This programme enhances real-time notifications by grabbing specific video clips according to predefined standards. When considering wirelessly connected smart camera networks, this becomes quite important.

Object Detection

Using bounding boxes, object detection locates an object within a picture. To use YOLOv8 for detection, no suffix has to be added.

2.1.2 Automatic Unusual Activity Alerts

Provide a clear definition of odd activity within the parameters of your surveillance system. This could apply to any behaviour that is deemed odd, including overcrowding, abrupt movements, and loitering.

Develop or enhance your YOLO model to identify patterns or characteristics associated with atypical behaviours in addition to common objects. This could be creating new classes or changing ones that already

exist to include the behaviours you wish to identify.

Automatic Forensic Video Retrieval (AFVR)

In Automatic Forensic Video Retrieval (AFVR), pertinent information is often retrieved from video footage by means of computer vision techniques like object detection. In this situation, YOLO (You Only Look Once) can be a useful tool for object detection.

Define Forensic Objectives:

Clearly state what you want your forensic video retrieval system to accomplish. Choose whatever particular items or actions from the video clip you wish to identify and extract.

Learn to Love Your Life Objects (YOLO): Develop or improve your YOLO model to find items that are important for forensic examination. This could apply to people, cars, or certain objects of interest. Make use of a labelled dataset containing instances of these things in different contexts.

2.2 Situation Awareness

Situation awareness is being aware of your surroundings, looking for trends, and using the information at hand to make judgement calls. Incorporating YOLO (You Only Look Once) into a situation awareness system helps improve real-time comprehension of visual data.

Describe the circumstances:

- Give a clear description of the conditions and factors that make up the situation you wish to be aware of. This could refer to certain items, pursuits, or circumstances in a particular setting.

Develop YOLO for Useful Objects:

- Develop or improve your YOLO model to find objects that are pertinent to the given scenario. Make sure the things and situations you want to be aware of are represented in your dataset.

Real-Time Object Detection:

- Use YOLO to apply real-time object detection to recorded or live video sources.

3. VIDEO SURVEILLANCE SYSTEM

In a smart city, a vehicle surveillance system (VSS) is made to keep an eye on the surrounding area through a variety of devices, including cellphone cameras, unmanned aerial vehicles (UAVs), dash cams from cars, and public or private CCTVs. end camera devices, an edge computing system that processes and analyses video data to provide users with a fast response, a cloud computing system that has sufficient processing power and storage to allow users to analyse the video data in-depth using deep learning, and a blockchain system that facilitates safe storage and anonymous user consensus on the video data are all components of the VSS architecture. The technologies and apparatuses associated with the VSS differ based on its scope and objectives. For instance, edge computing is commonly utilised for environmental monitoring and in situations where real-time transmission and prompt action are essential, like in hospitals. Deep learning techniques are used to analyse the footage acquired by roadside cameras in a traffic monitoring system. Furthermore, as a first step in the early fire detection process, a color-based deep learning classification algorithm is applied to ascertain whether the incoming films' frames include red colour. Since cameras are now widely installed in many urban areas, it is simple to gather visual data from the surrounding region for use in smart city applications. Two categories exist for monitoring devices: both stationary and mobile monitoring equipment. To continuously monitor defined regions, fixed monitoring equipment are usually placed on streets, at traffic crossings, and inside and outside of buildings. Moving monitoring devices, on the other hand, are made to move freely and keep an eye on places that are not readily visible. classifies monitoring systems for the smart city VSS according to how mobile they are; some cameras, like CCTVs, are stationary, while other cameras, like those on mobile phones and cars, are mobile.

4. PROPOSED WORK

One of the issues that many security systems have is their inability to carry out specific tasks without a driver monitoring the system's progress automatically. Humans are not meant to work for twenty-four hours a day. There is a limit to how long mortal drivers can drive while they are awake. People cannot stay awake for extended periods of time without being distracted, and they will always have to sleep. Because of this, the multi-camera videotape system has special features that might have a big impact on security assiduity. The

technology is very specialised, and it can assist a lot of individuals in overcoming security issues that they encounter on a daily basis. It's crucial to take these factors into account while choosing what kinds of security systems are required to be borrowed.

SCOPE

Intelligent video processing competencies are possible within the scopes of prevention, detection, and intervention that have resulted in the development of genuine and reliable video surveillance systems. Advanced video-based surveillance can be broadly defined as an intelligent video processing approach intended to facilitate effective video analysis for forensic investigations and to help security professionals by delivering dependable real-time alerts. The requirements for creating a solid and dependable video surveillance system are covered in this chapter. Additionally, the many kinds of cameras needed for various environmental scenarios—such as monitoring in both interior and outdoor spaces—are covered. Various modelling approaches are needed to build an effective surveillance system in different lighting scenarios.

4.1.3 Software and Hardware Requirements

Hardware Specification: -

- Processor – i5 and above (64-bit OS).
- Memory – 4GB RAM
- Hard Disk – 64 GB
- Input devices – Keyboard, Mouse.

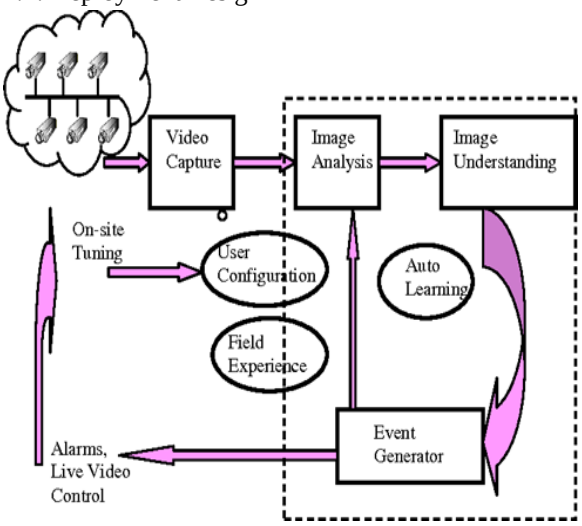
Software Specification: -

- Python: Language in which code is written
- CMake: For compiling openCV
- Visual Studio Code: For building openCV and darknet code
- Nvidia GPU Driver: For faster GPU performance
- OpenCV: For working on images/videos in python

4.1 Problem Statement

Conventional video surveillance systems are vulnerable to environmental fluctuations, such as shifts in light, water-induced background agitation, and light reflections. Utilising automatic video analysis technology is essential to creating intelligent surveillance systems that can assist drivers in identifying and responding to implicit hazards. The Smart Video Surveillance System (SVSS) has the ability to analyse videotape-based grounded objects.

4.2. Deployment Design



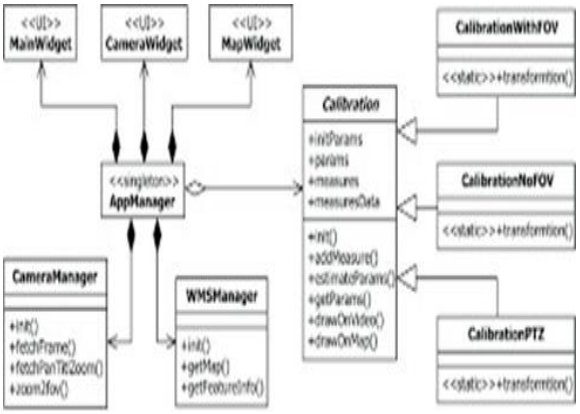


Fig..4.3.1 Class Diagram

4.3 Implementation Plan:

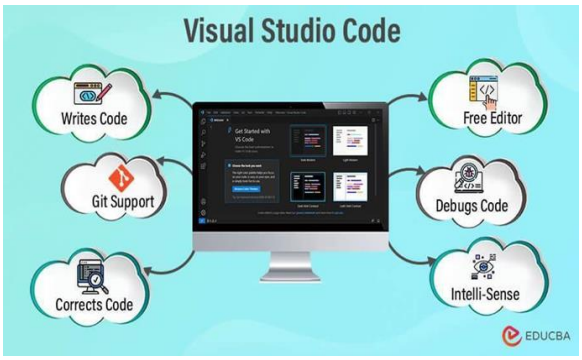
4.3.1 ENVIRONMENTAL SETUP

Microsoft created Visual Studio Code, usually known as VS Code, which is a source-code editor available for Windows, Linux, and macOS. Support for debugging, snippets, code rewriting, intelligent code completion, and embedded Git are among the features. Users have the ability to modify the theme, keyboard shortcuts, preferences, and install functional extensions. According to 86,544 respondents in the Stack Overflow 2023 Developer Survey, Visual Studio Code is the most widely used developer environment tool, with 73.71% of respondents saying they use it. Additionally, according to the poll, those learning to code use Visual Studio Code more frequently than expert developers (78% vs. 74%).

Microsoft first revealed Visual Studio Code on April 29, 2015, during the Build conference. A sneak peek build was made available. not too long after that. Visual Studio Code's source code was publicly available on GitHub and distributed under the MIT Licence on November 18, 2015. Support for extensions was also declared. The public preview version of Visual Studio Code ended on April 14, 2016, and it was made available online. While Microsoft's binary releases of Visual Studio Code are freeware and contain proprietary code, the majority of the program's source code is available on GitHub under the permissive MIT Licence. There is a community distribution called VSCodium that offers binaries with an MIT licence. Numerous programming languages, such as C, C#, C++, Fortran, Go, Java, JavaScript, Node.js, Python, Rust, and Julia, can be used with Visual Studio Code, a source-code editor.

Key Features:

- Cross-platform compatibility: Operates without a hitch on Linux, macOS, and Windows.
- Open-source and free: Anyone can use and contribute to it.
- Several programming languages are supported, including PHP, Go, Java, Python, JavaScript, TypeScript, C++, C#, and many more. Extensions are available for even more languages.
- As you type, intelligent code completion, or IntelliSense, suggests variables, methods, and keywords to help you write code more quickly and precisely.
- Debugging: Step through code, set breakpoints, and examine variables with the built-in debugging tools for different languages.
- Git integration: Git version control functionality, such as committing, branching, and merging code, are built-in.
- Add-ons: a sizable marketplace full of extensions that let you personalise VS Code with features like language support, code snippets, themes, and more.
- Customisable: Tailor the editor's settings, keyboard shortcuts, and appearance to your tastes.



5. CONCLUSION

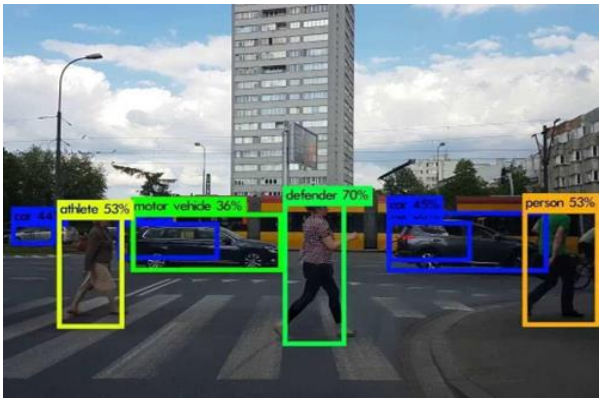


Fig. 4.4.3 Final-Stage-2

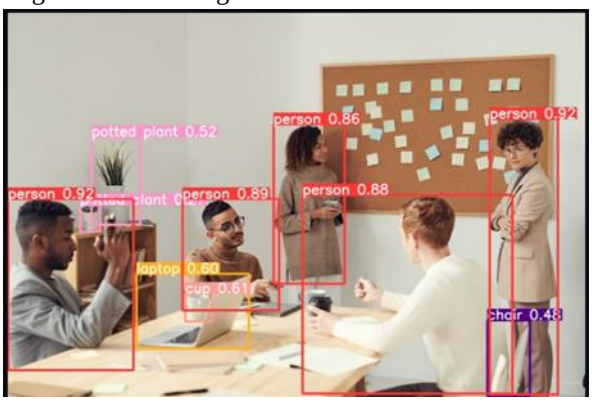


Fig. 4.4.4 Final-Stage-3

Identifying and localising certain objects in a video stream is made possible through object classification, which is essential to VSSs. Object classification is used to classify and group items into predetermined classifications, such as people, cars, plants, animals, and environmental elements in a city area, once they have been recognised in a video stream. By grouping items into distinct categories, we are able to collect the data needed for additional processing and analysis.

Several processes make up the object classification process, including feature extraction, object

recognition, picture pre-processing, and image capture. Video compression, noise reduction, and picture enhancement methods are used in the image pre-processing step. In order to find patterns in the image, feature extraction is done. Then, to identify the object of interest, object recognition compares the derived characteristics to a database of known things. Several criteria, including motion, form, colour, and texture, as outlined in the next subsection, can be used to realise object recognition if feature extraction is completed prior to object recognition.

VSSs frequently employ the motion-based object detection technique to identify and track objects like cars, humans, and animals based on their motion characteristics.

Analysing variations in motion within a video is the process of motion-based object detection.

order to recognise areas or noteworthy items.

Outstanding results were obtained with intelligent video surveillance that used the YOLO object detection method. when models were trained using the accessible information. Because the system was using optimised algorithms, it was able to identify objects and classes, detect motion, and perform with more accuracy. With confounded predictions, the neural networks and algorithms that were employed fared well. The creation of the GUI made it possible for users to use features more realistically and efficiently. Each module also improved overall user experience by operating well.

6. FEATURES OF VIDEO SURVEILLANCE SYSTEM IN SMART CITIES

The essential features of a VSS for edge/cloud computing, deep learning techniques, and camera surveillance of urban environments. These features primarily consist of object detection, HAR, and surveillance video analysis.

There are several motion detection algorithms that can be applied, including optical flow, background subtraction, and frame differencing. These algorithms work well in environments that change quickly and require less computing power.

The authors presented the W4 algorithm with frame differencing for recognising moving cars and pedestrians in a noisy setting with a complicated background. This algorithm uses the W4 algorithm and frame differencing separately to calculate the difference between frames. The results of each technique are then combined using a logical OR operation. In order to identify the final object in the combined result and filter out the noise, morphological operations with associated component labelling are finally used.

REFERENCE

[1] Ben Dickson. (2020) What are Convolutional Neural Networks.https://www.experfy.com/blog/ai-ml/what-areconvoluti onal-neural-networks-cnn/

[2] The structure of an artificial neuron, the basic component of artificial neural networks. Source: Wikipedia.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi (2016). You Only Look Once: Unified, Real-Time Object Detection.arXiv:1506.02640v5 [cs.CV]

[4] Chengji Liu, Yufan Tao, Jiawei Liang, Kai Li1, Yihang Chen Object (2018). Detection Based on YOLO Network

[5] Wenbo Lan, Jianwu Dang, Yangping Wang, Song Wang (2018). Pedestrian Detection Based on YOLO Network Model

[6] Rumin Zhang, Yifeng Yang (2018). An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera.

[7] Joseph Redmon, Ali Farhadi (2018). YOLOv3: An Incremental Improvement, arXiv:1804.02767

[8] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár (2015) MicrosoftCOCO: Common Objects in Context