IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# INTELLIGENT SURVILLENCE SYSTEM

[1] Ms. Chinju Poulose, [2]Ashlin Francis Pereira, [3]Sushith K S, [4]Sreekumar C M, [5]Yadhukrishna K Suresh

[1]Assistant Professor, Department of Computer Science And Engineering, Universal Engineering College, Vallivattom , Thrissur, India.

[2,3,4,5] B. Tech Student, Department of Computer Science And Engineering, Universal Engineering College, Vallivattom , Thrissur, India.

***Abstract:*** In contemporary modern-day world, safety and safety are fundamental worries. To be economically robust, a rustic have to provide a secure and secure environment for investors and tourists. Closed Circuit television (CCTV) cameras, however, are used for surveillance and monitoring sports. A large number of surveillance cameras are to be had in numerous locations, but all of them simplest file footage. To analyse these videos, a large quantity of manpower is required, that's constantly unwanted because of time and labour waste. As a end result, having a system that detects crime in real time and signals the user is advantageous. This paper proposes a device for automatically detecting crime from surveillance digital camera pictures. The device has been pre-trained to hit upon crimes in real time and makes each online and offline alerts, that's useful in decreasing the opportunity of criminals escaping and quickly arrest them. If the lacking person enters the premises, the system detects her or him.

***Index Terms* – Face recognition, weapon detection, CNN**

## I. INTRODUCTION

The crime rate across the globe has increased mainly because of the frequent use of handheld weapons during violent activity. For a country to progress, the law-and-order situation must be in control. High incidents were recorded in past few years with the use of harmful weapons in public areas and in most of this incidents there is a very hard role of suspects like most wanted criminals. Also the number of missing cases are more in today world. The amount of missing person increases due to personal issues, depression, ecnomic un-stability etc. So it is relevent to identify these suspects and harmful weapons from the public places. This paper presents a system for detecting crimes and locating missing people in real time using surveillance digital camera photographs and informing the user about the incidence. For the machine's success, image processing techniques and machine learning are employed. Humans currently operate the current CCTV tracking system. Its an huge time consuming process. They take a seat at a desk in front of a bank of display screens and conduct ongoing non-specific analysis of live video feeds to determine an event. The proposed system simplifies this task by reducing the human effort on this. The system is pretrained to detect weapons and also to recognize the suspects like criminals and missing persons. So whenever a suspect or weapon is detected in surveillance permesis the system alerts the user about the scenario. Thus the user can take necessary action for it.

## II. RELEATED WORK

Here we discuss the various studies based on automation on survillance system.

The paper[1] implies a framework that provides a secure place using CCTV footage as a source to detect harmful weapons by applying the state of the art open-source deep learning algorithms. This paper implements binary classification assuming pistol class as the reference class and relevant confusion objects to reduce false positives and false negatives. It uses two approaches sliding window/classification and region proposal/object detection. Some of the algorithms used are VGG16, Inception-V3, Inception-ResnetV2, SSDMobileNetV1, Faster-RCNN Inception-ResnetV2 (FRIRv2), YOLOv3, and YOLOv4. and from test results its shown that YOLOv4 gives good results in terms of Precision and recall count and f1-score.

The paper[2] affords a singular algorithm for detection of certain varieties of unusual events. The set of rules is based on multiple local monitors which collect low-stage information. each neighborhood reveal produces an alert if its modern-day dimension is unusual and those alerts are integrated to a very last choice regarding the lifestyles of an uncommon event. This set of rules satisfies a set of requirements which are crucial for a success deployment of any big-scale surveillance gadget. specifically, it requires a minimum setup (taking only some mins) and is completely computerized afterwards. instead of seeking to song gadgets, proposed algorithm monitors low-level measurements in a fixed of fixed spatial positions. The authors present consequences on films from 8 cameras in 5 special websites. except for the eating hall and bus terminal websites, where the authors recorded with the very own digital camera, all videos had been received by direct recording from website online-established surveillance cameras

The paper[3] implies, as an lively studies topic in pc vision, visual surveillance in dynamic scenes attempts to hit upon, recognize and song positive gadgets from image sequences, and extra typically to apprehend and describe object behaviour. The stipulations for powerful computerized surveillance the usage of a unmarried camera include the subsequent levels: modelling of environments, detection of movement, class of shifting items, monitoring, knowledge and outline of behaviour, and human identity. The authors offer specific discussions on future research guidelines in visual surveillance, e.g., occlusion coping with, mixture of two-dimensional (2-D) tracking and three-D monitoring, aggregate of motion analysis and biometrics, anomaly detection and behaviour prediction, behaviour know-how and nature language description, content material-based retrieval of surveillance movies, fusion of information from a couple of sensors, and far off surveillance

In paper [4] , writer advise a fully unsupervised dynamic sparse coding method for detecting uncommon activities in motion pictures based on online sparse reconstructablity of question indicators from an atomically found out event dictionary, which bureaucracy a sparse coding bases. primarily based on an instinct that ordinary occasions in a video are more likely to be reconstructable from an event dictionary, whereas unusual activities are not, proposed algorithm employs a principled convex optimization formulation that lets in both a sparse reconstruction code, and an online dictionary to be at the same time inferred and updated. Proposed set of rules is absolutely unsupervised, making no earlier assumptions of what unusual events may additionally appear like and the settings of the cameras. Detection of irregular visual patterns in images and in video sequences is useful for a variety of tasks.

In this paper [5] , writer deal with the trouble of detecting irregularities in visual facts, e.g., detecting suspicious behaviours in video sequences, or figuring out salient styles in photographs . This paper [4] display programs of this approach to figuring out saliency in pics and video, for detecting suspicious behaviours and for automatic visual inspection for high-quality guarantee Detecting suspicious behaviours or uncommon gadgets is important for surveillance and tracking The composition method is applied as an efficient inference set of rules in a probabilistic graphical model, which contains for small spatio-temporal deformations between the query and the database

In paper [6], The authors introduces novel structural assumptions on the joint distributions to account for spatial and temporal locality of anomalies. The empirical composite scoring and rating scheme asymptotically converges to the most efficient selection rule for maximizing detection energy problem to false alarm constraints. Sparse decomposition for each spatio-temporal scale can be viewed as a characteristic vector that feeds into the local KNN block. this is due to the fact Gn(·) as defined inside the preceding section combines facts over neighborhood neighbourhoods of a information sample and the ranking function produces a composite rating for an entire random area.

The authors of paper[7] introduces novel structural assumptions on the joint distributions to account for spatial and temporal locality of anomalies. The empirical composite scoring and rating scheme asymptotically converges to the most efficient selection rule for maximizing detection energy problem to false alarm constraints. Sparse decomposition for each spatio-temporal scale can be viewed as a characteristic vector that feeds into the local KNN block. this is due to the fact Gn(·) as defined inside the preceding section combines facts over neighborhood neighbourhoods of a information sample and the ranking function produces a composite rating for an entire random area.

According to the paper[8] the volume of internet videos has been rising at an exponential price, necessitating a big increase inside the need for easier video looking. critical visual concepts tend to emerge repeatedly throughout a collection of movies with the same topic, and the frequency of visual co-occurrence may be used as a proxy to measure. The sparsity of co-taking place patterns provides an additional impediment to video co-summarization: hundreds to heaps of pictures may be seen in a single video; but, there are normally only some pictures that appear in a couple of recordings. The authors offer a completely unique Maximal Biclique coming across (MBF) method to cope with this problem, which formulates the problem as finding entire bipartite subgraphs that maximise total visual co-occurrence within a bipartite network. The authors provide a completely unique Maximal Biclique coming across (MBF) approach to address this hassle, which formulates the mission as finding complete bipartite subgraphs that maximise total visual co-incidence inside a bipartite graphical representation of images and motion pictures.

## III. PROPOSED SYSTEM

The proposed system can automatically detect the weapons and suspects who enter the survelliance area and alert the user. This system can reduce the effort and labor that was needed in existing system.

### A. METHODOLOGY:

The proposed system have three modules. The first module is the face detection. In this module the face is located in the captured frame in order to detect the suspects. The system determines the faces in frame by using the common face pattern which will have the structure of eyes, nose, ears, chin etc. In next step the system finds the coordinates of the detected frame. It find out the x-y coordinate of pixels in the face location. The OpenCV provides functions to detect the face in python. After getting the locations of faces in frame by the first module, the second module is used to recognized faces. For this this module uses a dataset that consist of pictures of suspects or person that should be recognized by the system. For fast computation of the system, these data are processed in a numerical form. So this module converts the images in database to 128 dimensional encodings values. That is from each picture in database the face location is determined and a 128-dimensional encoding is found for each detected face. This encoding value will be different for each person. Once encodings for all images in database is found, these values are compared with the face detected in the live captured image. It uses the Euclidean distance formula to find the distance between them. And the encodings that give minimum value during comparison is the person detected in the live frame. Thus the information about corresponding person can taken from database and alert the user. The third module is the weapon detection. Here the system is pretrained with the sample images of weapons like pistol, knife, etc and a Deep Neural Network is created to detect the weapons. So if a weapon is detected the system alert the user.

The system alert the user when a suspect or weapon detected in the frame. The system generates two type of message and one is sent to user registered email id and other to the system itself.

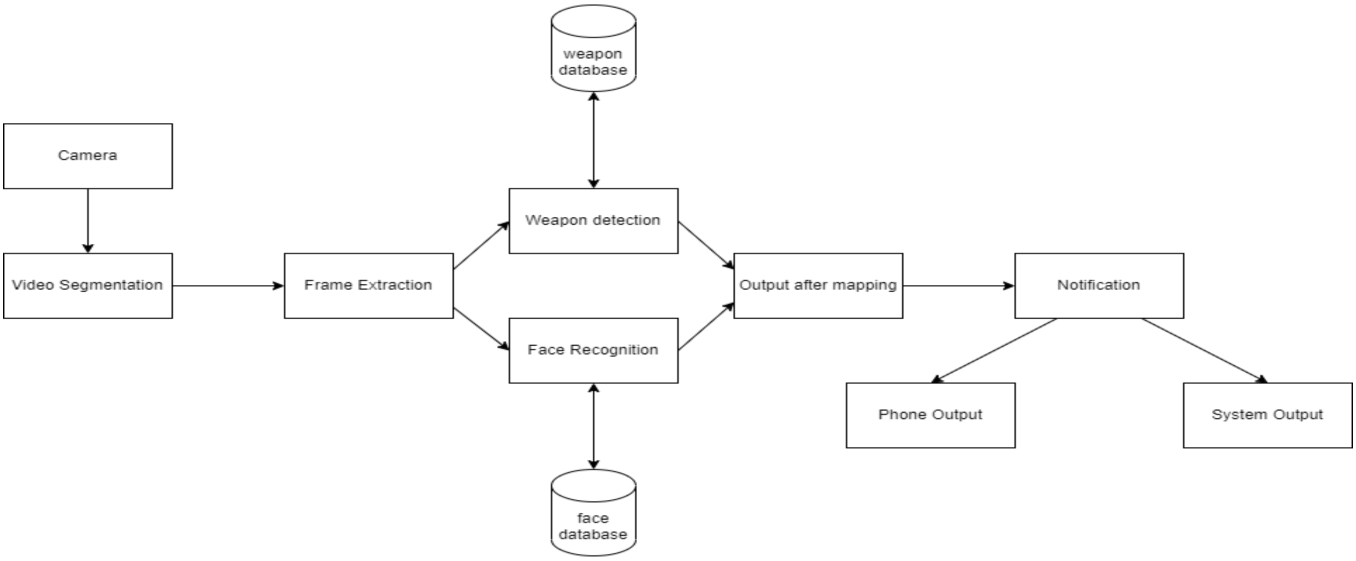


Fig.1: Block Diagram of Intelligent survelliance system

## B. FACE DETECTION AND FACE RECOGNITION:

The system uses the python as backend and in python there are many functions that helps to perform particular functionality. The python libiaries such as OpenCV, Face recognition, numpy etc helps to simplify the working of our system.For image processing the opencv is used and  for mathematical operations numpy libiary is used. Face recognition libiary provides functions to find the data like encodings ,locations, distance etc that are needed to detect and recognize the person from an image.

The below figures shows the struture of database used for face recognition in this system. The system uses one or more images of person that should be recognized and store them in subfolders with their names in a main directory. The system find the 128 dimeshional encodings for each image and this encoding is compared with the encodings found for the faces found from frames extracted from live feed and comparison is based on one to many and the minimum argument among them shows that the person is recognised by the system and finally system alerts the user.

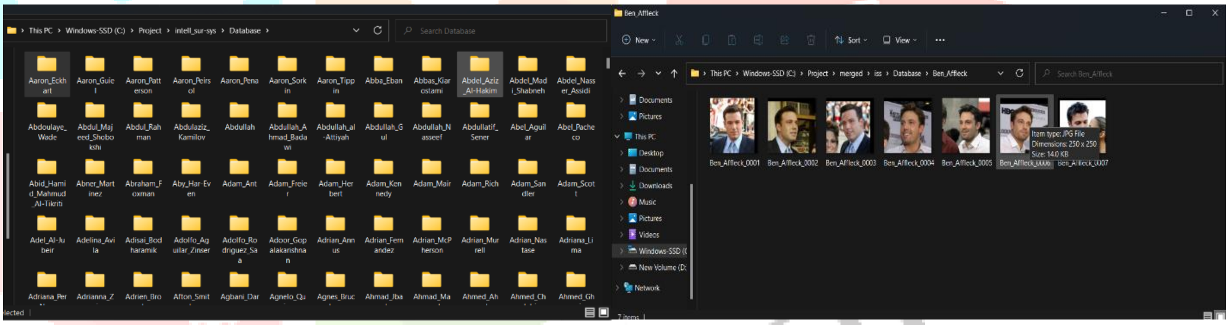
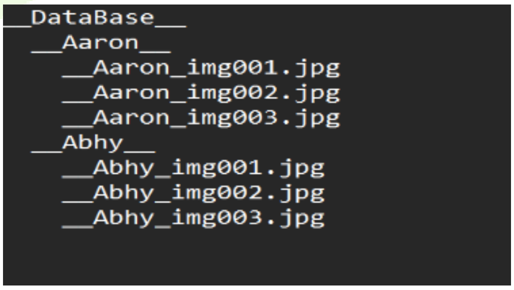


Fig 2: subfolders in database of face and sample image files in each subfolder



```
__DataBase__
    __Aaron__
        __Aaron_img001.jpg
        __Aaron_img002.jpg
        __Aaron_img003.jpg
    __Abhy__
        __Abhy_img001.jpg
        __Abhy_img002.jpg
        __Abhy_img003.jpg
```

Fig 3: File struture in face database

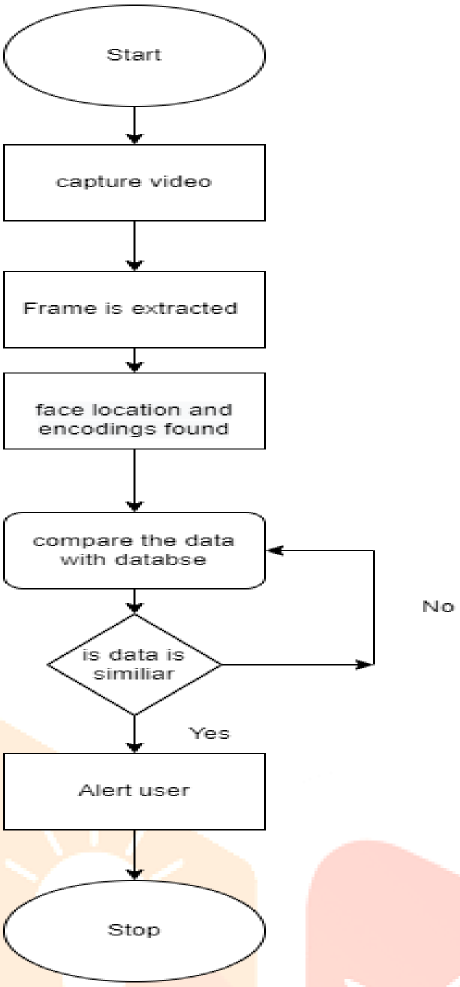
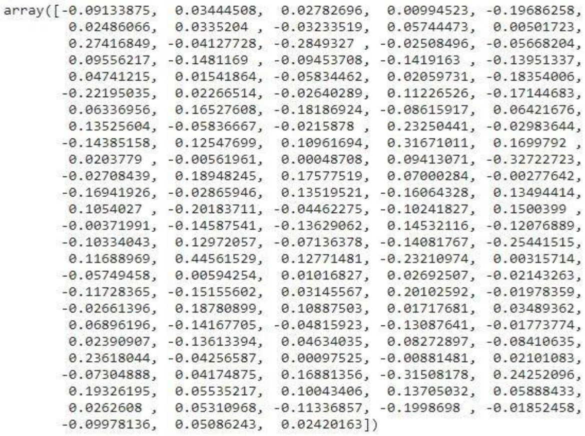


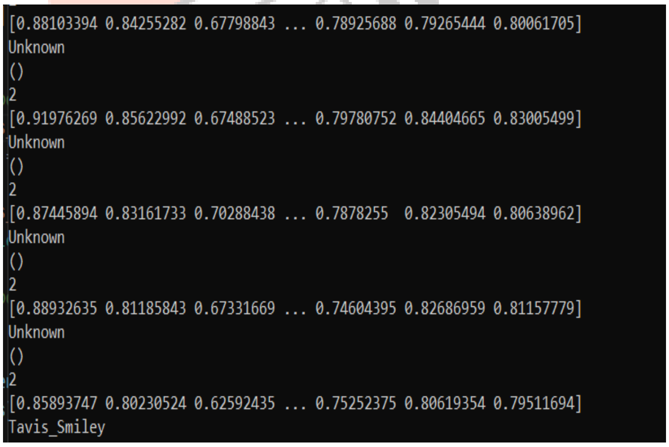
Fig 4: Flowchart of face recognition

The above shown flow chart shows the work flow of the face recognition module.



```
array([-0.09133875,  0.03444508,  0.02782696,  0.00994523, -0.19686258,
        0.02486066,  0.0335204 , -0.03233519,  0.05744473,  0.00501723,
        0.27416849, -0.04127728, -0.2849327 , -0.02508496, -0.05668204,
        0.09556217, -0.1481169 , -0.09453708, -0.1419163 , -0.13951337,
        0.04741215,  0.01541864, -0.05834462,  0.02059731, -0.18354006,
       -0.22195035,  0.02266514, -0.02640289,  0.11226526, -0.17144683,
        0.06336956,  0.16527608, -0.18186924, -0.08615917,  0.06421676,
        0.13525604, -0.05836667, -0.0215878 ,  0.23250441, -0.02983644,
       -0.14385158,  0.12547699,  0.10961694,  0.31671011,  0.1699792 ,
        0.0203779 , -0.00561961,  0.00048708,  0.09413071, -0.32722723,
       -0.02708439,  0.18948245,  0.17577519,  0.07000284, -0.00277642,
       -0.16941926, -0.02865946,  0.13519521, -0.16064328,  0.13494414,
        0.1054027 , -0.20183711, -0.04462275, -0.10241827,  0.1500399 ,
       -0.00371991, -0.14587541, -0.13629062,  0.14532116, -0.12076889,
       -0.10334043,  0.12972057, -0.07136378, -0.14081767, -0.25441515,
        0.11688969,  0.44561529,  0.12771481, -0.23210974,  0.00315714,
       -0.05749458,  0.00594254,  0.01016827,  0.02692507, -0.02143263,
       -0.11728365, -0.15155602,  0.03145567,  0.20102592, -0.01978359,
       -0.02661396,  0.18780899,  0.10887503,  0.01717681,  0.03489362,
        0.06896196, -0.14167705, -0.04815923, -0.13087641, -0.01773774,
        0.02390907, -0.13613394,  0.04634035,  0.08272897, -0.08410635,
        0.23618044, -0.04256587,  0.00097525, -0.00881481,  0.02101083,
       -0.07304888,  0.04174875,  0.16881356, -0.31508178,  0.24252096,
        0.19326195,  0.05535217,  0.10043406,  0.13705032,  0.05888433,
        0.0262608 ,  0.05310968, -0.11336857, -0.1998698 , -0.01852458,
       -0.09978136,  0.05086243,  0.02420163])
```

Fig 5:128-D encodings



```
[0.88103394 0.84255282 0.67798843 ... 0.78925688 0.79265444 0.80061705]
Unknown
()
2
[0.91976269 0.85622992 0.67488523 ... 0.79780752 0.84404665 0.83005499]
Unknown
()
2
[0.87445894 0.83161733 0.70288438 ... 0.7878255  0.82305494 0.80638962]
Unknown
()
2
[0.88932635 0.81185843 0.67331669 ... 0.74604395 0.82686959 0.81157779]
Unknown
()
2
[0.85893747 0.80230524 0.62592435 ... 0.75252375 0.80619354 0.79511694]
Tavis_Smiley
```

Fig 6: comparison values.

The above figures shows the 128 dimensional encodings found by the system and the values found during comparison and the value which is minimum among all is taken and corresponding file that gives that minimum argument is recognize as the data of detected person. Fig 6 shows that an face is recognized (Travis_smilery).

For faster detection and recognition the system first finds the encodings for all person an store it as a file. This can reduce the work burden of the system. That is if this is not done the system have to do encoding for database on each and every running of the system. So the proposed system creates a file in format of list with all encodings of all images and labels ie the name of person that should be recognized.

### C. WEAPON DETECTION:

This is the second most important module of our proposed system. For weapon detection we creates a pretrained model that detects the weapon in real time. For this the first step is to create a database of weapon image and classify them under specific class of weapons ie for example gun's images are classified under label Gun and knifes under Knife.

The next step is labelling and annotation. That is on each image, the coordinates of weapon is detected and bounding boxes are drawn. On each weapon it is labelled with corresponding class and dataset in yolo format is generated. The fig 8 shows the process of drawing bounding box around the images in dataset. This process is done for all images in the database and the locations and class labels for all weapons in dataset is found. Fig 9 shows the locations represented in yolo format as our system uses the yolo algorithm for weapon detection. The figure below shows the work flow of the weapon detection module.

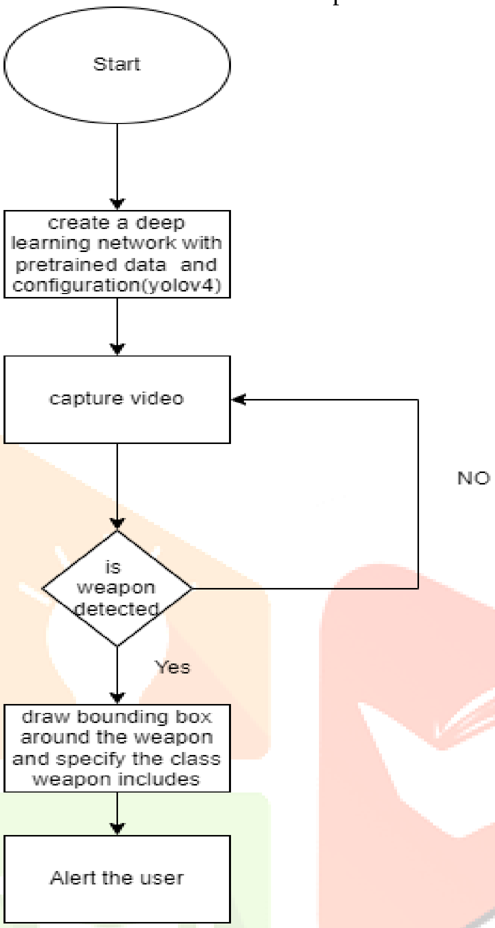

Start
create a deep learning network with pretrained data and configuration(yolov4)
capture video
is weapon detected
NO
Yes
draw bounding box around the weapon and specify the class weapon includes
Alert the user

Fig 7: Flowchart of weapon detection

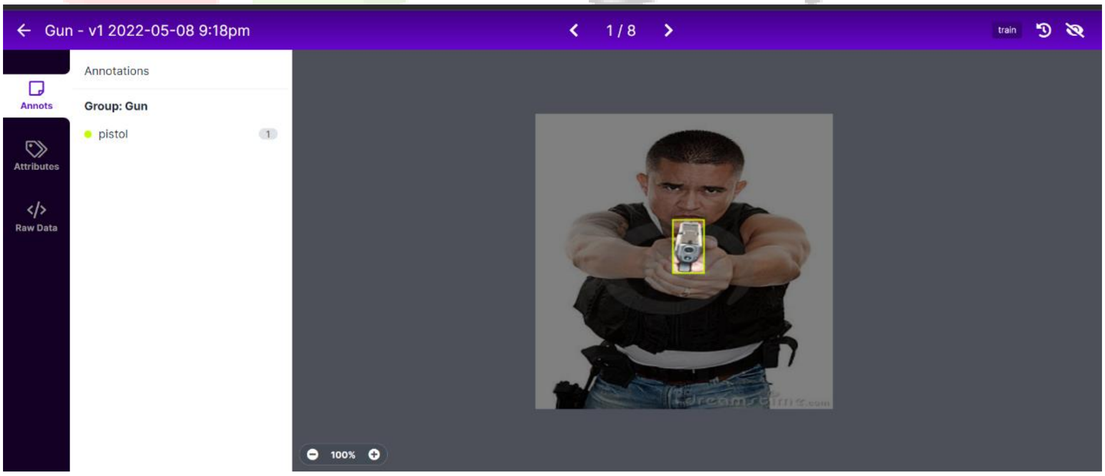
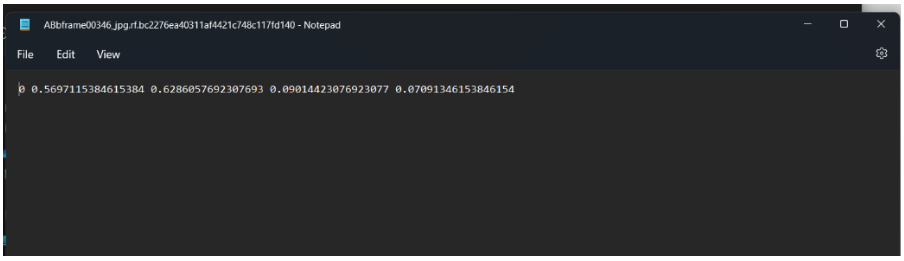


Fig 8: Image Annotation and Labelling


ABbframe00346_jpg.rf.bc2276ea40311af4421c748c117fd140 - Notepad
File
Edit
View

```
0 0.5697115384615384 0.6286057692307693 0.09014423076923077 0.07091346153846154
```

Fig 9: Bounding Box coordinates in Yolo format

After the process of annotation and labelling a dataset in format of yolo is obtained and the next step is to tarin a model. For this the dataset is divided into two one for testing and another one for validating the model. For training the model we used the darknet. Darknet have deep neural network arcitickture. Different approaches are used in the work for weapon classification and detection purpose but all have deep learning and CNN architecture behind them because of their state-of-the art performance.

There are several ways to generate region proposals, but the simplest way of generating them is Sliding window/Classification Models. In the method to the sliding window, a box or window is moved over a picture to select an area and use the object recognition model to identify each frame patch covered by the window. Yolo is for You Only Look Once, and it is a state-of-the-art method that operates on a real-time system based on deep learning to solve numerous Object y problems.

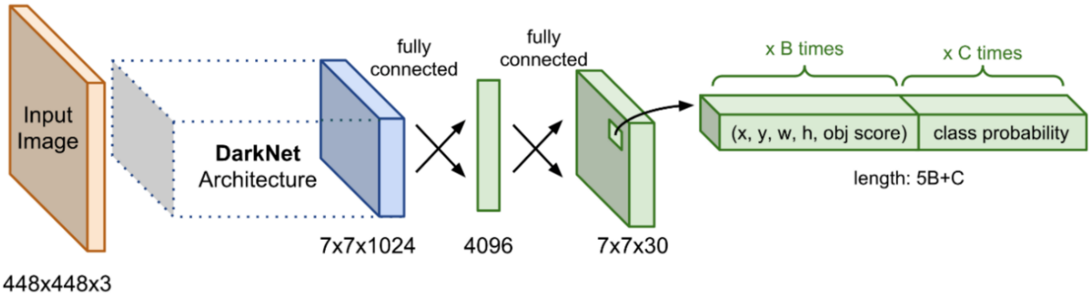


Fig 10: Yolo Architecture

The architecture contains the Input picture layers, as can be seen in the diagram above. And input depends upon the use case. Next layer is DarkNet Architecture. It tris an open-source neural network framework built with C and CUDA that includes YOLO for object recognition and tracking. Training is conducted using notebook software such as Google Colab, Juypter Notbook, and others. This creates an text document that consist of class and coordinates of bounding box that is x, y, width, height for each and every image in the dataset.

Next step in training of model is to generate classes.names file. This file consist of all class names in it in our case classes are knife and Gun. Next is to create train.txt and test.txt which contains file paths of our training images and testing images. This helps in testing and validating our dataset. If the loss on training set is high, it means the model is unfit and need to train longer. If the loss is low training set and high on test set, it means model is overfitting and need to add more data. The training iterates for several times. And its shows that after 2000 iterations there is no much change in precision and accuracy of the model.

This model generates a weights file and by using this file and configuration file used for training the yolo model we creates an neural network which can predict the classes. The model check the confidence value and the time taken to predict and thus determine the accuracy of the model.

After creating the network with custom config file and training weights file the system can detect the weapon whenever it appears in the detection frame.

The graph given below shows that the model doesn't gain any improvement in accuracy precision after 2000 iterations. The training is stopped when the CNN deployed the model with high accuracy in prediction. This data is validated before deployement by finding loss and accuracy in both train and test data.

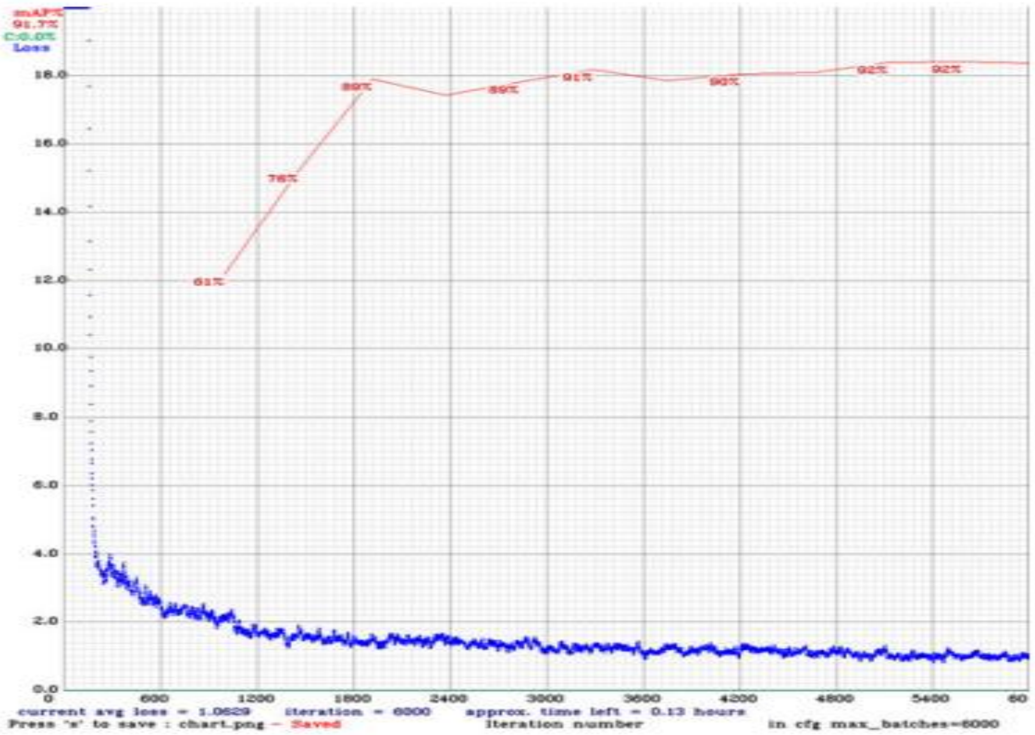


Fig 11: Graph : Loss Vs mAp

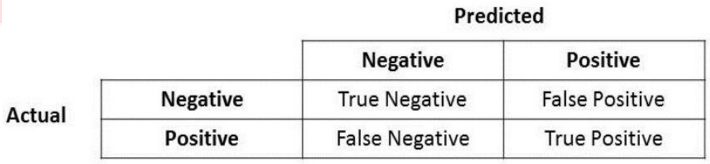
### D. ALERT SYSTEM

After the setting the face recognition module and weapon detection module ready the next step is to alert the user whenever this modules detects the event they trained for that is detect suspects or weapons. The system generates two type of alerts. One is sent to user email id and another one is pop up in UI screen of the user.

## IV. RESULTS AND DISCUSSIONS

The intelligent survelliance system is pretrained to detect the suspects and weapons that shown in the survelliance area. And whenever the system detects an suspect or weapons it alert the user via online and offline. This helps to prevent the crime before it happens or become worse.

The face recognition module can detect and recognize the faces in good accuracy. The accuracy can be increases by adding more than one picture of the person who should be recognized by the system. The use of dataset in numerical form increases the speed of the proposed system. The weapon detection module also shows a good accuracy in detection of weapons in real time. The accuracy of the system is determined by using the confusion matrix. With help of this the accuracy and precision can be found easily. The matrix consist of four components True positive, True negative, False positive, and False negative.



| | | Predicted | |
|---|---|---|---|
| | | Negative | Positive |
| Actual | Negative | True Negative | False Positive |
| | Positive | False Negative | True Positive |

Fig12 : Confusion Matrix

We used about 250+ images for guns and Knife and about 4000+ images for face recognition. A true positive occurs when the model correctly predicts the positive class. A true negative, on the other hand, is when the model predicts the negative class correctly. A false positive arises when the model incorrectly predicts the positive class. A false negative occurs when the model predicts the negative class incorrectly.

In face recognition the process of finding encodings of each image take about average of 1.5 minutes each. So for encoding an dataset of 4000+ images it took around 1 hour for the encoding. So once it found before staring the system and store as a file it reduces the work effort and time of proposed system.

The precision accuracy and f1 scores of pretrained model can be found with help of the confusion matrix. The forumals for finding the accuracy of the model are:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

On comparison of Accuracy and F1-score for VGG, Inceptionv3, InceptionResNetv2, SSDMobileNet, FasterRCNN InceptionResNetv2, Yolov3 and yolov4 the yolov4 gives an good result. For the model.

| SR NO | MODELS | IoU THRESOLOD= 50% | | |
|---|---|---|---|---|
| | | PRECISION | RECALL | F1-SCORE |
| 1 | SSDMobileNet | 62.7% | 60% | 59% |
| 2 | Yolov3 | 85.8% | 87% | 80% |
| 3 | FasterRCNN | 86.3% | 89% | 87% |
| 4 | yolov4 | 93% | 88% | 91% |

Table 1: Object detection accurcy on various models

The yolov4 model gives the best performance and system predicts the wepons and its clasess more accuratliy The figures shown below are the output from the proposed system.

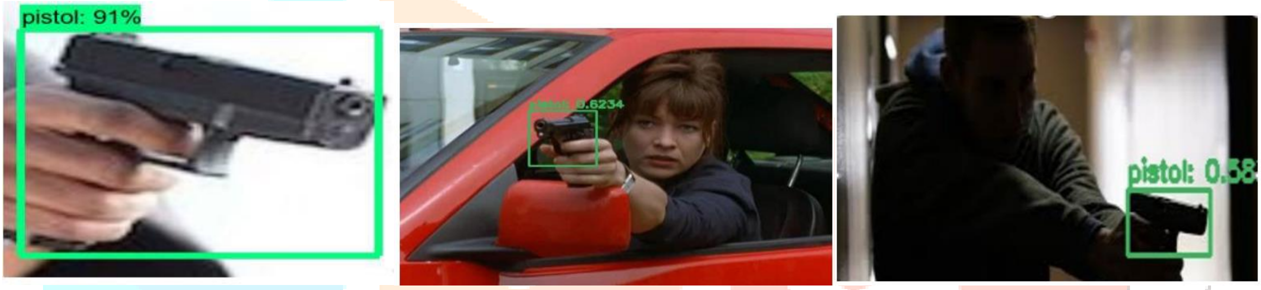


Fig 13: Weapon detection results

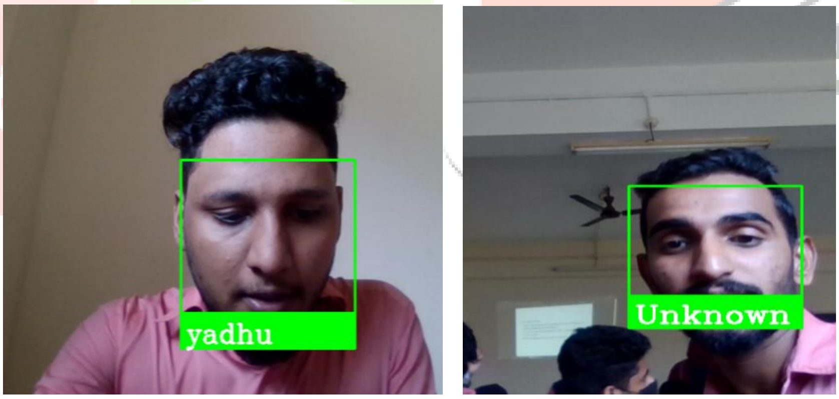
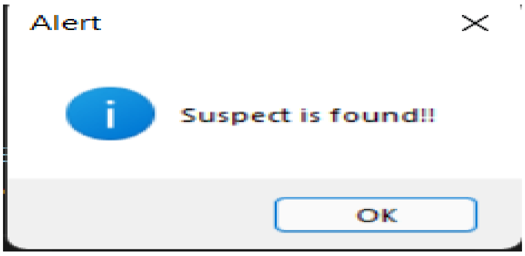
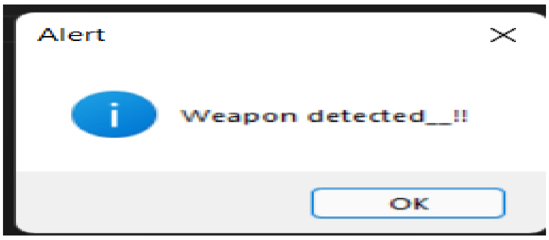


Fig 14: Face Recognition



Suspect is found!!

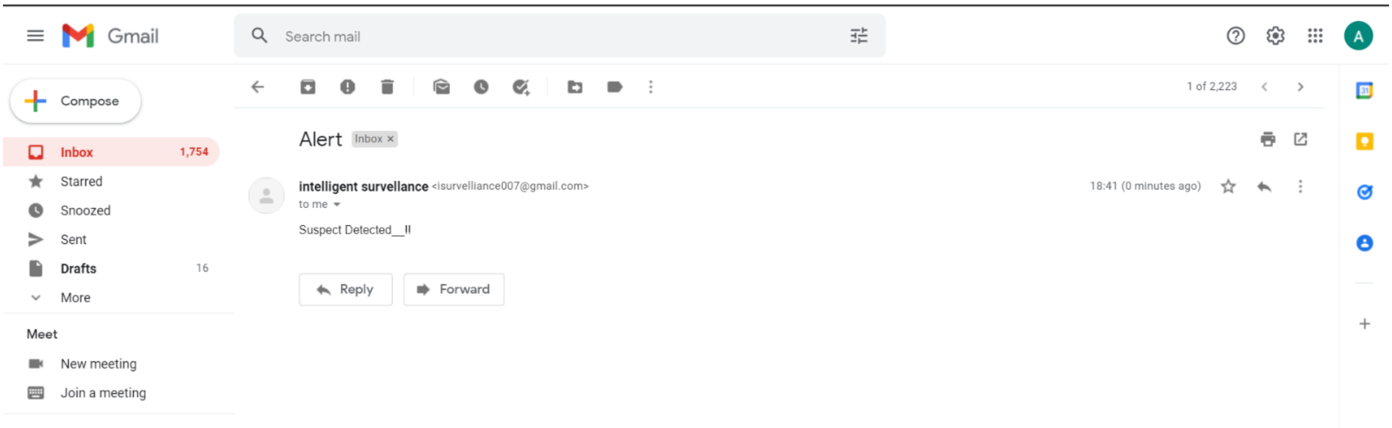
Fig 15(a): Offline Alert for face recognition



Alert

Weapon detected__!!

Fig 15(b): Offline Alert for weapon detection



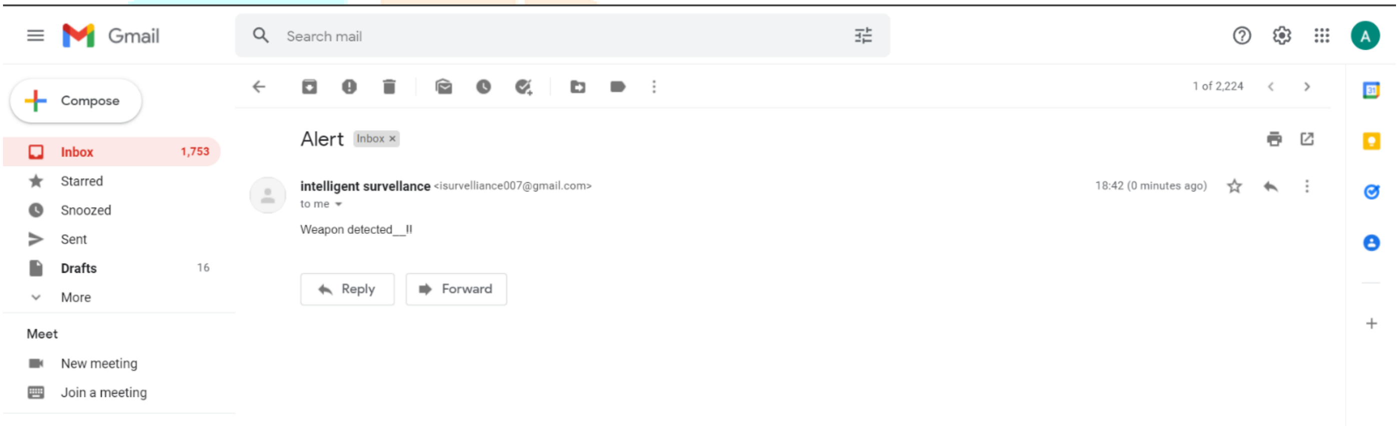
Alert Inbox ×

intelligent survellance <isurvelliance007@gmail.com>
to me

18:41 (0 minutes ago)

Suspect Detected__!!

Fig 16(a) Online Suspect Alert



Alert Inbox ×

intelligent survellance <isurvelliance007@gmail.com>
to me

18:42 (0 minutes ago)

Weapon detected__!!

Fig 16(b) Online Weapon Alert

## V. CONCLUSION AND FUTURE SCOPE

Intelligent surveillance system can gain a good attention due to the increasing demand on security and safety. The system is able to automatically analyze image, video or other type of surveillance data without human intervention. And also able to alert the user when an unusual event occurs. The system generates two type of alerts one is pop op as alert message on the computer itself and other one is sent to the user's registered email ID. The system helps to reduce the effort that was taken to monitor the survelliance footages. Also the system helps to prevent most of the crimes before it happens or become worse. The proposed system can be also implemented in existing survelliance system. That is most of the shops and private properties will have a CCTV survelliance system which record the events in real time. and if a unusual event occurred these footages can only used as proof of crime. But if we implement this proposed system , the system will detect the crime or suspects at the instance that happens or enter the permesis. It also alert the user in real time even the user is out of station. Thus the user can take necessary actions to prevent the crime at time it happens. The system can perform more better and faster on a GPU enabled device.

In future the system can detect the hidden weapons by using thermal cameras and Also system can be pretrained to detect the human action and predict weather a crime happens or not. With help of good efficient cameras the system can detect the event in long distance. The system can be re-modelled to alert the authorities like police department by sending alert message with the location to nearby police unit so the suspects cannot escape very easily.

## VI. ACKNOWLEDGEMENT

We have a great pleasure in publishing the paper paper titled “Intelligent Survillence System”. And also wish to record our indebtedness and thankfulness to all those who helped us to prepare it. First and foremost, we praise God Almighty for His providence and for being the project's guiding light. We are quite grateful to Dr. Jose K Jacob, UEC's Principal, for his invaluable guidance and encouragement. We are really grateful to Dr. Sreeraj R -HOD, CSE for his cooperation and assistance. We are especially thankful to our coordinator, guide and supervisor Ms. Chinju Poulose– Assistant Professor, UEC for her meticulous guidance and support that helped us in the completion of this project. Lastly, we would like to thank all faculties of Computer Science and Engineering department for the support and suggestions that helped us in the development of our project. We would also like to thank our colleagues who helped us directly or indirectly during this project.

## REFERENCES

**[1].** Muhammad Tahir Bhatti, Muhammad Gufran Khan , (Senior Member, Ieee), Masood Aslam, And Muhammad Junaid Fiaz "Weapon Detection in Real-Time CCTV Videos Using Deep Learning" IEEE Received January 13, 2021, accepted February 1, 2021, date of publication February 12, 2021, date of current version March 4, 2021

**[2].** Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D.: ”Robust real-time unusual event detection using multiple fixed-location monitors”. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(3), 555- 560 (2008)

**[3].** W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 34, no. 3, pp. 334–352, Aug 2004.

**[4].** B. Zhao, L. Fei-Fei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in CVPR, 2011, pp. 3313–3320.

**[5].** O. Boiman and M. Irani.” Detecting irregularities in images and in video”. In ICCV, pages 462–469, 2005

**[6].** Saligrama, V., and Chen, Z. 2012.”Video anomaly detection based on local statistical aggregates”. In CVPR, 2112–2119. IEEE.

**[7].** R. Mehran, A. Oyama, and M. Shah. ”Abnormal crowd behavior detection using social force model”. CVPR, 2009

**[8].** W.-S. Chu, Y. Song, and A. Jaimes, “Video Cosummarization: Video Summarization by Visual Cooccurrence,” in CVPR, 2015