PROJECT REPORT ON

SMART SURVEILLANCE SYSTEM FOR AUTISM CENTRES

in partial fulfillment of the requirements for the degree of Bachelor of Electronics and Communication Engineering.

Submitted by,

KARTHIKA G S (TKM17EC076)

KRISHNAJA T RAMESH (TKM17EC081)

NAVYA SUSAN JACOB (TKM17EC091)

SHAMINI P R (TKM17EC115)

Under the guidance of

Prof. SHAFI M N

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

TKMCE KOLLAM, KERALA, 691005

2020-2021

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM, KERALA, 691005

CERTIFICATE

This is to certify that the project report entitled **“SMART SURVEILLANCE SYSTEM FOR AUTISM CENTRES”** is a bonafide account of the project presented by **KARTHIKA G S (TKM17EC076), KRISHNAJA T RAMESH (TKM17EC081), NAVYA SUSAN JACOB (TKM17EC091) and SHAMINI P R (TKM17EC115)**in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Electronics and Communication of APJ Abdul Kalam Technological University during the academic year 2020-2021 under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Prof. SHAFI M N	**Prof. NISHANTH N**	**Prof.ABID HUSSAIN**
Project Guide	**Project Coordinator**	**HOD**

DECLARATION

We undersigned hereby declare that the project report “SMART SURVEILLANCE SYSTEM FOR AUTISM CENTRES ”, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under supervision of Prof. Shafi M N and Prof. Nishanth N. This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place: Kollam

Date: 13-07-2021

KARTHIKA GS	KRISHNAJA T RAMESH	NAVYA SUSAN JACOB	SHAMINI P R
(TKM17EC076)	(TKM17EC081)	(TKM17EC091)	(TKM17EC0115)

ACKNOWLEDGEMENT

It is with great pleasure and pride that we present this report before you. At this moment of triumph, let us thank all those who helped us in the successful completion of this report.

We pay our deep sense of gratitude to Dr. T A Shahul Hameed, Principal, for providing all the necessary facilities.

We are extremely grateful to Prof. Abid Hussain, Head of the Department, Department of ECE, for his constant encouragement and advice.

We are deeply thankful to Prof. Sajeena A, Associate Professor and Senior Advisor, Department of ECE, for all her support and encouragement.

We extend our sincere gratitude to our project coordinator Prof.Nishanth N, Assistant Professor, Department of ECE, for his guidance and suggestions

We gratefully acknowledge the exemplary encouragement, supervision, suggestions and inspired guidance offered by our guide Prof. Shafi M N, Assistant Professor, Department of ECE, TKMCE, in bringing this report to a successful completion.

We would like to extend our sincere & immense gratitude towards all the faculties of Department of Electronics and Communication Engineering and our friends who have helped us in this endeavor for their active guidance, help, cooperation & encouragement.

Last but not the least we thank the Almighty for his grace and care.

KARTHIKA G S

KRISHNAJA T RAMESH

NAVYA SUSAN JACOB

SHAMINI P R

ABSTRACT

The health of the soul and the body is one of the important issues requiring more attention.Autism is a developmental disorder characterized by difficulties with social interaction and communication, and by restricted and repetitive behavior. Parents often notice signs during the first three years of their child's life. These signs often develop gradually, though some autistic children experience regression in their communication and social skills after reaching developmental milestones at a normal pace.Technology is used in almost all areas of life to simplify tasks, save time and improve convenience. But perhaps one of the most valuable uses of technology is when it is used to help individuals with Autism Spectrum Disorder (ASD).

The objective of this project is to create a Smart Surveillance System which can monitor the inmates of Autism centers, Mental Asylums or Special Schools for autistic children where monitoring of the inmates is very challenging and requires careful attention.

The project implements the proposed system using a pre-trained YOLOv3 model and image processing for video analysis, event detection and is also capable of notifying the caretakers of emergency situations.The system monitors the inmates using cameras installed inside the institution building and its premises and gives alert to caretakers.

(1) when the inmates moves out of the institution compound

(2) when they possess harmful objects that can hurt themselves or other inmates.

The above system can aid the caretakers in centres or schools for ASD.

CONTENTS

LIST OF ABBREVIATIONS

AAP	American Academy of Pediatrics
ACM	Active Contour Models
API	Application Programming Interface
ASD	Autism Spectrum Disorder
CDC	Centre for disease control & prevention
CLI	Command Line Interface
CNN	Convolutional Neural Network
COCO	Common Objects in Context
GPU	Graphics Processing Unit
GSM	Global System for Mobile Communications
HOG	Histograms of Oriented Gradients
IOU	Intersection Over Union
IP	Internet Protocol
LSTM	Long Short Term Memory
mAP	mean Average Precision
RCNN	Region based Convolutional Neural Network
ResNet	Residual neural network
RPN	Region Proposal Network
SID	Security IDentifier
SMS	Short Message Service
SPP	Spatial Pyramid Pooling
SSD	Single Shot Multibox Detector
TPU	Tensor Processing Unit
VGG	Visual Geometry Group
YOLO	You Only Look Once
YOLOv3	You Only Look Once version 3

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It's the 21st century and we need to think smartly to make our life better, easier and secure, so instead of sitting at one place for longer why not carry the security in our pocket. Using security cameras allows a person to monitor his property. The majority of organizations and administrations are making use of such security cameras with the intention to save their business as well as property from terrorists and illegal entry. Nowadays, the security cameras have become much more advanced, reasonable, smaller and straightforward. Surveillance is seen as a measure of prevention and protection against security threats likely to happen in public spaces such as public squares and buildings, streets, hospitals, schools and other publicly accessible areas of the country.

1.1 MOTIVATION

"Children with autism are colorful – they are often very beautiful and, like the rainbow, they stand out."― Adele Devine

ASD refers to a spectrum of disorders with a range of manifestations that can occur on different degrees and in a variety of forms.To address the many challenges faced by children and adults with ASD, the role of sensing technologies becomes critical.

Centres for Disease Control & Prevention (CDC) estimates that approximately 1 in every 88 children are diagnosed with some level of autism, (one in 54 boys) a sharp jump from the previous numbers released in late 2009, and frighteningly distant rate from one in 10,000 cases seen in the1980s. Experts estimate that every 2-6 children out of every 1000 have Autism.

The prevalence rate of autism in India is 1 in 250 (figure may vary as many cases are not diagnosed)and currently 10 million people are suffering in India. The government only recognized the disorder in 2001, till 1980s, there were reports that Autism didn't exist in India (Dr. Vinod Kumar Goyal, TOI).

These alarming rates of increase call for attention by all the fields and clearly architecture has been ignoring the effect of the built environment in their development.

Studying the mind-set of autistic in itself is a very vast subject hence we focus on a few objectives to enhance security & safety in autism centres with the aid of image processing.

1.2 OBJECTIVES

Objective of this project is to create an Intelligent Surveillance System[4] which can monitor the inmates of Autism centers.
This system uses Machine Learning & Image Processing techniques for video analysis, event detection and is also capable of notifying the caretakers of emergency situations. The proposed system monitors the inmates using cameras installed inside the institution building and its premises and gives an alert to caretakers.
Outdoor camera checks for any inmate moving out of the institution compound
Indoor camera monitor if an inmate possesses harmful objects that can hurt themselves or others.

1.3 AUTISM SPECTRUM DISORDER

1.3.1 Screening and Diagnosis

Diagnosing autism spectrum disorder (ASD) can be difficult because there is no medical test, like a blood test, to diagnose the disorder. Doctors look at the child's developmental history and behavior to make a diagnosis.

ASD can sometimes be detected at 18 months or younger. By age 2, a diagnosis by an experienced professional can be considered very reliable. However, many children do not receive a final diagnosis until much older. Some people are not diagnosed until they are adolescents or adults. This delay means that children with ASD might not get the early help they need.

Early signs of ASD can include, but are not limited to

- Avoiding eye contact,

- Having little interest in other children or caretakers,
- Limited display of language (for example, having fewer words than peers or difficulty with use of words for communication), or
- Getting upset by minor changes in routine.

As children with ASD become adolescents and young adults, they might have difficulties developing and maintaining friendships, communicating with peers and adults, or understanding what behaviors are expected in school or on the job. They may also come to the attention of healthcare providers because they have co-occurring conditions such as attention-deficit/hyperactivity disorder, obsessive compulsive disorder, anxiety or depression, or conduct disorder.

Monitoring, screening, evaluating, and diagnosing children with ASD[12] as early as possible is important to make sure children receive the services and support they need to reach their full potential. There are several steps in this process.

i. Developmental Monitoring

Developmental monitoring observes how a child grows and changes over time and whether the child meets the typical developmental milestones in playing, learning, speaking, behaving, and moving. Parents, grandparents, early childhood providers, and other caregivers can participate in developmental monitoring.

ii. Developmental Screening

Developmental screening takes a closer look at how your child is developing. Your child will get a brief test, or you will complete a questionnaire about your child. The tools used for developmental and behavioral screening are formal questionnaires or checklists based on research that ask questions about a child's development, including language, movement, thinking, behavior, and emotions. Developmental screening can be done by a doctor or nurse, but also by other professionals in healthcare, community, or school settings.

Developmental screening is more formal than developmental monitoring and normally done less often than developmental monitoring. Your child should be screened if you or your doctor have a concern. However, developmental screening is a regular part of some of the well-child visits for all children even if there is not a known concern.

The American Academy of Pediatrics (AAP) recommends developmental and behavioral screening for all children during regular well-child visits at these ages:

- 9 months
- 18 months
- 30 months

In addition, AAP recommends that all children be screened specifically for ASD during regular well-child doctor visits at:

- 18 months
- 24 months
- Additional screening might be needed if a child is at high risk for ASD (e.g., having a sister, brother or other family member with an ASD) or if behaviors sometimes associated with ASD are present.

If a child is at higher risk for developmental problems due to preterm birth, low birthweight, environmental risks like lead exposure, or other factors, healthcare providers may also discuss additional screening. If a child has an existing long-lasting health problem or a diagnosed condition, the child should have developmental monitoring and screening in all areas of development, just like those without special healthcare needs.

iii. Comprehensive Developmental Evaluation

A brief test using a screening tool does not provide a diagnosis, but it indicates if a child is on the right development track or if a specialist should take a closer look. If the screening tool identifies an area of concern, a formal developmental evaluation may be needed. This formal evaluation is a more in-depth look at a child's development, usually done by a trained specialist, such as a developmental pediatrician, child psychologist, speech-language pathologist, occupational therapist, or other specialist. The specialist may observe the child, give the child a structured test, ask the parents or caregivers questions, or ask them to fill out questionnaires. The results of this formal evaluation determines whether a child needs special treatments or early intervention services or both.

1.3.2 Treatment and Intervention Services

Currently, no treatment has been shown to cure ASD, but several interventions have been developed and studied for use with young children. These interventions may reduce symptoms, improve cognitive ability and daily living skills, and maximize the ability of the child to function and participate in the community .

The differences in how ASD affects each person means that people with ASD have unique strengths and challenges in social communication, behavior, and cognitive ability. Therefore, treatment plans are usually multidisciplinary, may involve parent-mediated interventions, and target the child's individual needs.

Behavioral intervention strategies have focused on social communication skill development—particularly at young ages when the child would naturally be gaining these skills—and reduction of restricted interests and repetitive and challenging behaviors. For some children, occupational and speech therapy may be helpful, as could social skills training and medication in older children. The best treatment or intervention can vary depending on an individual's age, strengths, challenges, and differences .

Types of Treatments

There are many types of treatments available. These include applied behavior analysis, social skills training, occupational therapy, physical therapy, sensory integration therapy, and the use of assistive technology.

The types of treatments generally can be broken down into the following categories:

1. Behavior and Communication Approaches
2. Dietary Approaches
3. Medication

CHAPTER 2

LITERATURE REVIEW

[1] You Only Look Once: Unified, Real-Time Object Detection by Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi introduces YOLO, a unified model for object detection.

This model is simple to construct and can be trained directly on full images. Unlike classifier-based approaches, YOLO is trained on a loss function that directly corresponds to detection performance and the entire model is trained jointly. Fast YOLO is the fastest general-purpose object detector in the literature and YOLO pushes the state-of-the-art in real-time object detection. YOLO also generalizes well to new domains making it ideal for applications that rely on fast, robust object detection. The base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems,YOLO makes more localization errors but is less likely to predict false positives in the background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

[2] SSD Object Detection Model Based on Multi-Frequency Feature Theory by Jinling Li, Qingshan Hou , Jinsheng Xing, and Jianguo Ju,

This paper presents an improved SSD multi-object detection model that is proposed to improve the accuracy and real-time performance of the traditional Single Shot Multibox Detector (SSD) object detection model. Firstly, aiming at the defect of weak correlation between prediction object score and positioning accuracy in the traditional SSD model, the improved model enhanced the correlation between the two by adding Intersection Over Union(IoU) prediction loss branch. Secondly, in order to reduce the spatial redundancy of the traditional SSD model, a multi-frequency feature component convolution module is designed ,which greatly reduces the calculation overhead and hardware overhead of the traditional model. Finally, in order to accelerate the convergence speed of the improved model, the Adaptive and Momental Bound

(AdaMod) optimizer is introduced to modify the adaptive learning rate of the improved model which is too large in the training process. Experimental results show that the improved model has stronger detection capabilities, better overall detection results, and improved detection accuracy and real-time detection.

[3] Real Time Object Detection and Tracking Using Deep Learning and OpenCV by Chandan G, Ayush Jain, Harsh Jain, Mohana

This paper presents various model training methods. Deep learning has gained a tremendous influence on how the world is adapting to Artificial Intelligence since the past few years. Some of the popular object detection algorithms are Region-based Convolutional Neural Networks (RCNN), FasterRCNN, Single Shot Detector (SSD) and You Only Look Once (YOLO). Amongst these, Faster-RCNN and SSD have better accuracy, while YOLO performs better when speed is given preference over accuracy. Deep learning combines SSD and Mobile Nets to perform efficient implementation of detection and tracking. This algorithm performs efficient object detection while not compromising on the performance. Main Objective of SSD algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained and can be further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance. This real time analysis of the ecosystem can yield great results by enabling security, order and utility for any enterprise.

[4] Video Surveillance in Mental Health Facilities: Is it Ethical? by Tali Stolovy, Yuval Melamed, Arnon Afek

The paper introduces the security measures taken in the hospitals. Patient safety is a growing issue which can be improved with the usage of high-end centralised surveillance systems allowing the staff to focus more on treating health issues rather than keeping a watchful eye on potential incidents. Video surveillance is a tool for managing safety and security within public spaces. In mental health facilities, the major benefit of video surveillance is that it enables 24 hour monitoring of patients, which has the potential to reduce violent and aggressive behavior. The major disadvantage is that such observation is by nature intrusive. It diminishes privacy, a factor of huge importance for psychiatric inpatients. Thus, an ongoing debate has developed following the increasing use of cameras in this setting. This article presents the experience of a medium-large academic state hospital that uses video surveillance, and explores the various ethical and administrative aspects of video surveillance in mental health facilities.

[5] S. Xu and K. Hung, Development of an AI-based System for Automatic Detection and Recognition of Weapons in Surveillance Videos.

Focuses on security cameras and video surveillance systems for ensuring safety and security of the general public. The lack of manpower in the security sector and limited performance of humans may result in undetected dangers or delay in detecting threats, posing risks for the public. In response, various parties have developed real-time and automated solutions for identifying risks based on surveillance videos. The aim of this work is to develop a low-cost, efficient, and artificial intelligence-based solution for the real-time detection and recognition of weapons in surveillance videos under different scenarios. The system was developed based on Tensorflow and preliminarily tested with a 294-second video which showed 7 weapons within 5 categories, including handgun, shotgun, automatic rifle, sniper rifle, and submachine gun. At the intersection over union (IoU) value of 0.50 and 0.75, the system achieved a precision of 0.8524 and 0.7006, respectively.

[6] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks by Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun details Faster RCNN.

State-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations. Advances like SPPnet and Fast R-CNN have reduced the running time of these detection networks, exposing region proposal computation as a bottleneck. In this work, they introduce a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully-convolutional network that simultaneously predicts object bounds and objectness scores at each position. RPNs are trained end-to-end to generate high quality region proposals, which are used by Fast R-CNN for detection. With a simple alternating optimization, RPN and Fast R-CNN can be trained to share convolutional features. For the very deep VGG-16 model , the detection system has a frame rate of 5fps (including all steps) on a GPU, while achieving state-of-the-art object detection accuracy on PASCAL VOC 2007 (73.2% mAP) and 2012 (70.4% mAP) using 300 proposals per image.

[7] Detection of Anomalous Behavioural Patterns In University Environment Using CNN-LSTM by D. O. Esan, P. A. Owolawi and C. Tu

This is an experimental paper on surveillance systems.Inaccurate detection of anomalous behavioural patterns due to false detection (false alarm) errors is the main challenge that affects the performance of surveillance systems. Although several interesting techniques have been reported in literature for the detection of anomalous behaviour using semi-supervised and supervised techniques, these still require improvement in terms of accuracy and minimizing false detection errors in crowded environments. This paper presents a performance analysis of the convolutional neural network with long short-term memory (CNN-LSTM) in surveillance systems. The image features are extracted from the image frame sequences using the CNN, while the LSTM uses the gate mechanism to keep vital information for remembrance. The results are compared with existing detection models, including the mixture of probabilistic principal analysis, motion deep net, social force and dictionary-based models. Experiments are done on the University of California San Diego dataset using the proposed anomalous behavioural pattern detection system. The results obtained show that the proposed system outperforms the others mentioned with 86% accuracy.

[8]. SMS Notifications using Raspberry pi by P.Priyanka, T.Sivapriya, Sd.Laheen, V.Lakshmi Lahari Assistant Professor, Department of ECE, Audisankara institute of Technology, Nellore ,India.

This paper explains a raspberry pi controlled sms –update –notification(sun) system ,raspberry pi is a credit card sized single board computer with Arm11 microprocessor short message service (sms) is one of the cheapest and best way for sending a message from mobile .The basic idea of sun system is to notify updates to the people in a working environment from authorized persons at any where just by sms. This task is accomplished by conjunction of a GSM module with raspberry pi. The working procedure of this sun system is that only the authorized person can send sms from anywhere to the sim in the GSM module. This message will be read and displayed through a website on a monitor by raspberry pi.

[9] Object Tracking by Color and Active Contour Models Segmentation. by A. S. Silva, F. M. Q. Severgnini, M. L. Oliveira, V. M. S. Mendes and Z. M. A. Peixoto, Member, IEEE

This work presents a novel approach for an object recognition and tracking system, here known as SRCCA (Software de Reconhecimento de Cor e Contorno Ativo or Active Contour and Color Recognizing Software) by using color recognition in the HSV color space and active contour models (ACM). The main innovations proposed make use of the association of distinct techniques employed for object tracking and detection, as well as the elaboration of an open source software through the use of C++ programming language and OpenCV functions. To enhance the tracking efficiency for targets in motion, the video capturing device was controlled by two servo motors, enabling the target to be followed, in case it moves in the x- or y-axis. The final system was tested and presented robust and efficient tracking and contour segmentation capabilities for targets in motion, regardless of shape and size.

CHAPTER 3

THEORY

3.1 Video Processing

A video signal is a sequence of 2D images, captured onto the image plane of a video camera from a dynamic 3D scene. Basically in a color scene, the color value obtained at any point in a video frame record is the reflected light at a particular 3D point in the observed scene.

Video communication using video phones, video streaming, and video broadcasting[9,10] has become more and more attractive with the fast pace of development in the wireless industry in the present era. In order to transmit such rich multimedia content, it is desirable to occupy less bandwidth, the size of the original video signal must be reduced by some compression technique, without degrading video quality or data loss. To get rid of such hurdles, video compression techniques provide efficient solutions to represent video data in a more compact and robust manner so that the storage and transmission issues of video can be realized in a cost effective way. Digital video has become a necessary part of everyday life. Apart from compression it is well known that video enhancement as an alive topic in video processing has received much concentration in recent times.

Processing a video means performing operations on the video frame by frame. Frames are particular instances of the video in a single point of time. There may be multiple frames in a single second. Operations like adaptive threshold, smoothing, edge detection etc may be done while processing of video..

3.1.1 Steps Involved:

1. Video analysis step (to understand the video semantics)
2. Video image analysis step (to assess image relevance)
3. Manual identification of scene changes (to identify the target of the identification)

3.2 YOLO (YOU ONLY LOOK ONCE)

[1]YOLO is an abbreviation for the term 'You Only Look Once'. This is an algorithm that uses neural networks to detect and recognize various objects in a picture (in real-time). Object detection in YOLO is done as a regression problem and provides the class probabilities of the detected images. As the name suggests, the algorithm requires only a single forward propagation through a neural network to detect objects.

It works in three techniques:

1. Residual blocks,
2. Bounding box regression,
3. Intersection Over Union (IOU)

First, the image is divided into various grids. Each grid has a dimension of S x S. YOLO uses a single bounding box regression to predict the height, width, center, and class of objects. Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap. YOLO uses IOU to provide an output box that surrounds the objects perfectly. Each grid cell is responsible for predicting the bounding boxes and their confidence scores. The IOU is equal to 1 if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equal to the real box. The Architecture of YOLO has 24 convolutional layers and 2 fully connected layers.YOLO takes an input image and resizes it to 448×448 pixels. The image further goes through the convolutional network and gives output in the form of a 7×7×30 tensor. Tensor gives the information about 1) coordinates of the bounding box's rectangle and 2) Probability distribution over all classes the system is trained for. Thresholding these confidence scores (probability) eliminates class labels scoring lesser than 30%.

Compared to conventional methods of object detection like RCNN, faster RCNN,etc YOLO has certain advantages.

1) Rather than using a two-step method for classification and localization of an object, YOLO applies a single CNN for both classification and localization of the object.

2) YOLO can process images at about 40-90 FPS, so it is quite fast . This means streaming video can be processed in real-time, with negligible latency in a few milliseconds. The

architecture of YOLO makes it extremely fast. When compared with R-CNN, it is 1000 times faster and 100 times faster than fast R-CNN[6].

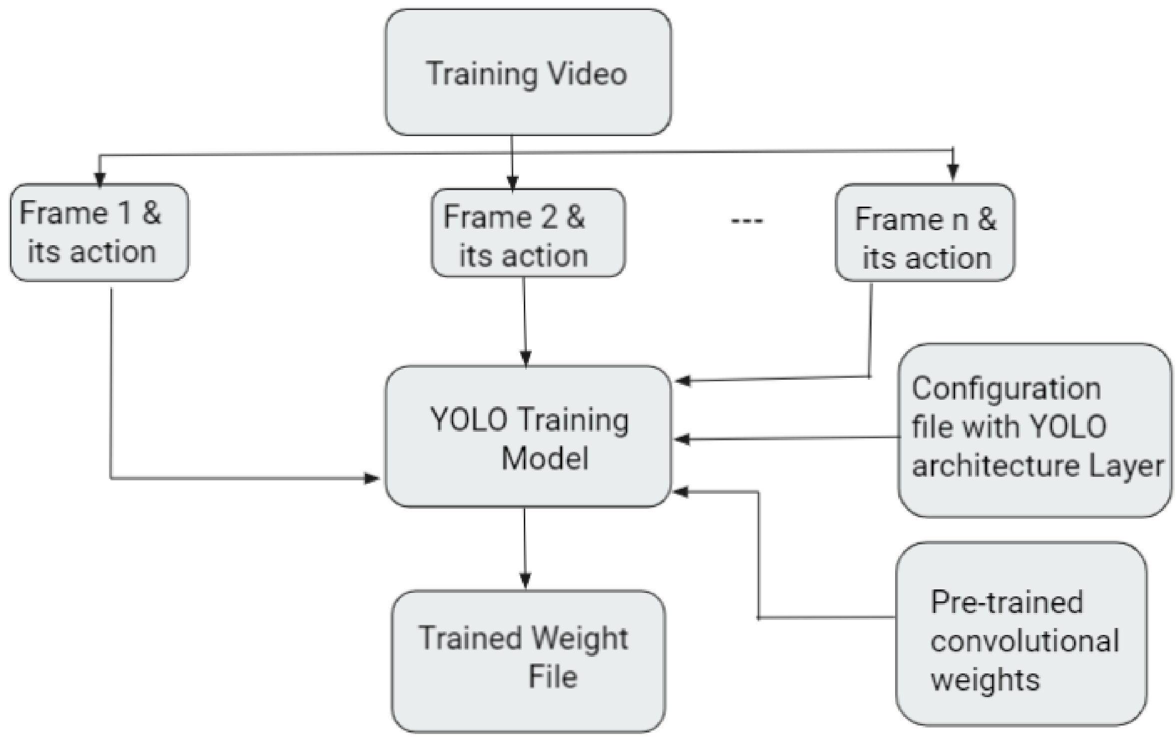


Figure 3.2 : Yolo block diagram

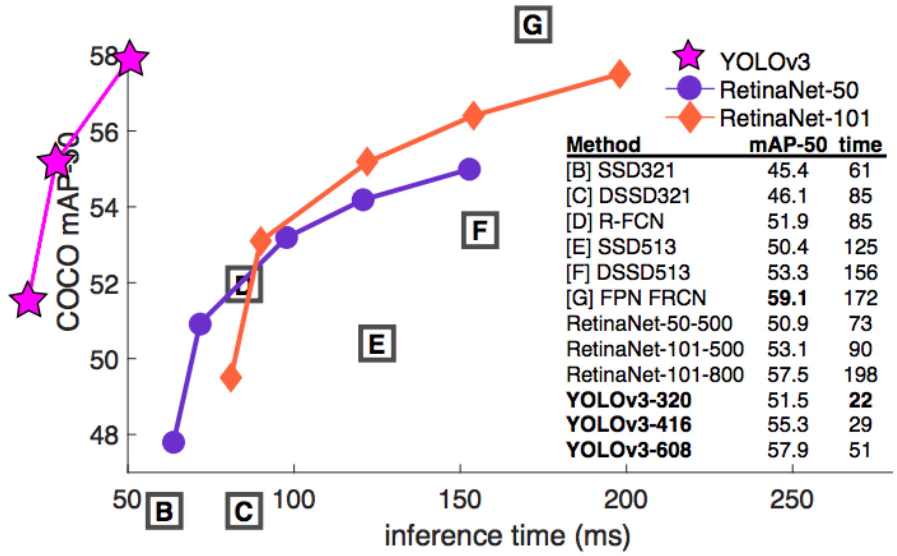
3.3 YOLOv3

The YOLOv3 algorithm[13] first separates an image into a grid. Each grid cell predicts some number of boundary boxes or anchor boxes around objects that score highly with predefined classes. Each boundary box has a respective confidence score of how accurate it assumes that prediction should be and detects only one object per bounding box. The boundary boxes are generated by clustering the dimensions of the ground truth boxes from the original dataset to find the most common shapes and sizes. Unlike systems like R-CNN and Fast R-CNN, YOLO is trained to do classification and bounding box regression at the same time.

Differences between YOLOv3 and older versions are in terms of speed, precision, and specificity of classes. YOLOv2 was using Darknet-19 as its backbone feature extractor, while YOLOv3 now uses Darknet-53. Darknet-53 has 53 convolutional layers instead of the previous 19, making it more powerful than Darknet-19 and more efficient than competing backbones. YOLOv3 is fast and accurate in terms of mAP(mean average precision) and intersection over union (IOU) values as well. It runs significantly faster than other detection methods with comparable performance.

The precision for small objects in YOLOv2 was incomparable to other algorithms because of how inaccurate YOLO[1] was at detecting small objects. With an AP of 5.0, it pales compared to other algorithms like RetinaNet (21.8) or SSD513 (10.2). The higher the AP, the more accurate it is for that variable. YOLOv3 increased the AP for small objects by 13.3, which is a massive advance from YOLOv2. However, the AP for all objects (small, medium, large) is still less than RetinaNet.

The new YOLOv3 uses independent logistic classifiers and binary cross-entropy loss for the class predictions during training. YOLOv3 uses a multilabel approach which allows classes to be more specific and be multiple for individual bounding boxes. Meanwhile, YOLOv2 used a softmax, a mathematical function that converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the relative scale of each value in the vector. Using a softmax makes it so that each bounding box can only belong to one class. The accuracy of detecting objects with YOLOv3 can be made equal to RetinaNet by having a larger dataset, making it an ideal option for models that can be trained with large datasets. But YOLOv3 may not be ideal to use with niche models where large datasets can be hard to obtain.



Method	mAP-50	time
[B] SSD321	45.4	61
[C] DSSD321	46.1	85
[D] R-FCN	51.9	85
[E] SSD513	50.4	125
[F] DSSD513	53.3	156
[G] FPN FRCN	**59.1**	172
RetinaNet-50-500	50.9	73
RetinaNet-101-500	53.1	90
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	**22**
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

Figure 3.3 Comparison of Algorithms

CHAPTER 4

TECHNOLOGICAL STACK

4.1 DESCRIPTION OF SOFTWARE AND PLATFORM

The program code is written in Python language and the platforms used are PyCharm IDE and Google Colab. Python's standard library is very extensive, offering a wide range of facilities. Our project uses several computer vision libraries like opencv, numpy, tensorflow, keras, imutils, Matplotlib for detections and visualisations.

4.1.1 OpenCV

OpenCV is a popular and open-source computer vision library that is focussed on real-time applications. The library has a modular structure and includes several hundreds of computer vision algorithms. OpenCV includes a number of modules including image processing, video analysis, 2D feature framework, object detection,camera calibration, 3D reconstruction and more.

4.1.2 Imutils

Imutils is a computer vision package that includes a series of OpenCV and convenience functions to make basic image processing functions such as translation, rotation, resizing, skeletonisation, displaying Matplotlib images, sorting contours, detecting edges, among others quite easy.

4.1.3 Matplotlib

Matplotlib is a comprehensive visualisation library for creating static, animated and interactive visualisations in Python. The library can perform various functions like developing publication-quality plots, export and embed to a number of file formats and interactive environments, and more.

4.1.4 NumPy

NumPy provides support for large, multi-dimensional arrays. Images can be expressed as multi-dimensional arrays.. Representing images as NumPy arrays is not only computational and resource efficient, but many other image processing and machine learning libraries use NumPy array representations as well. NumPy's built-in high-level mathematical functions can quickly perform numerical analysis on an image.

4.1.5 Keras

Keras is a high-level neural networks library that is capable of running on top of either TensorFlow or Theano. The library was developed with a focus on enabling fast experimentation. This deep learning library provides several features, including support for both convolutional networks and recurrent networks, allowing easy and fast prototyping, among others.

4.1.6 Tensorflow

TensorFlow is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. Tensorflow is a symbolic math library based on dataflow and differentiable programming.

4.1.7 Google Colab

Colaboratory is a Google Research product that is a hosted Jupyter Notebook service where python code can be written and executed through the browser and is specifically designed for machine learning, data analysis and education and gives free access to computing resources like GPU and TPU. Colab notebooks are stored in Google Drive, or can be loaded from GitHub. Code is executed in a virtual machine private to the user's account. Virtual machines are deleted when idle for a while, and have a maximum lifetime enforced by the Colab service.

Resources available in Colab vary over time to accommodate fluctuations in demand, as well as to accommodate overall growth and other factors.The GPUs available in Colab often include Ny. vidia K80s, T4s, P4s and P100s. Resources in Colab are prioritized for users who have recently used less resources, in order to prevent the monopolization of limited resources. Google Colab provides pre-installed machine learning libraries such as Pandas, NumPy, MatplotlibKeras,TensorFlow, and PyTorch. It runs on cloud and offers collaboration with multiple users thus making works easier to access, share and modify.

4.1.8 TWILIO (Alert generation)

A. To notify the caretakers in case of emergency we used a platform called Twilio .
B. Twilio allows users to make and receive phone calls, send and receive text messages, and perform other communication functions using its web service APIs.
C. With just a few lines of code,Python code can send SMS messages with Twilio.

Steps involved

1. Sign up for Twilio and get first SMS-enabled Twilio phone number
2. Install the Twilio CLI
3. Import Twilio Helper Library in the program
4. Update program with unique Account SID , Auth Token (generated while creating Twilio account) message body and sender/receiver number.

4.1.9 ImageAI 2.1.6

An open-source python library to build applications and systems with self-contained Deep Learning and Computer Vision capabilities using simple and few lines of code. ImageAI currently has a PyTorch backend. ImageAI provides abstracted and convenient implementations of Computer Vision technologies and supports a list of state-of-the-art Machine Learning algorithms for image prediction, custom image prediction, object detection, video detection, video object tracking and image predictions training. It currently supports image prediction and training using 4 different Machine Learning algorithms-MobileNetV2, ResNet50, InceptionV3 and DenseNet121, trained on the ImageNet-1000 dataset. Object detection, video detection and object tracking is done using RetinaNet, YOLOv3[13] and TinyYOLOv3 trained on COCO dataset. Using ImageAI, custom models can also be trained for performing detection and recognition of new objects. ImageAI provides very convenient and powerful methods to perform object detection in images & extract each object from the image. With ImageAI, detection tasks, analysis on videos or live-video feeds from device cameras and IP cameras can be performed using Video Object Detection and Tracking and Video analysis functions. ImageAI provides classes and methods to train a new model that can be used to perform prediction on custom objects (to detect any object of interest by providing the images, the annotations and training with ImageAI.) Training of custom models is using SqueezeNet, ResNet50, InceptionV3 and DenseNet.

CHAPTER 5

PROPOSED SYSTEM

The project implements the proposed system using a pre-trained YOLOv3 model and image processing for video analysis, event detection and is also capable of notifying the caretakers, The system can aid the caretakers in centres or schools for ASD. The proposed system consists of an indoor camera & an outdoor camera.

1. Indoor camera performs
 i)Monitoring inmates
 ii)Harmful object detection like knives and scissors
2. Outdoor camera monitors if any inmate is trying to move out of the compound and alert the caretaker.

5.1 OUTDOOR CAMERA

5.1.1 Block diagram

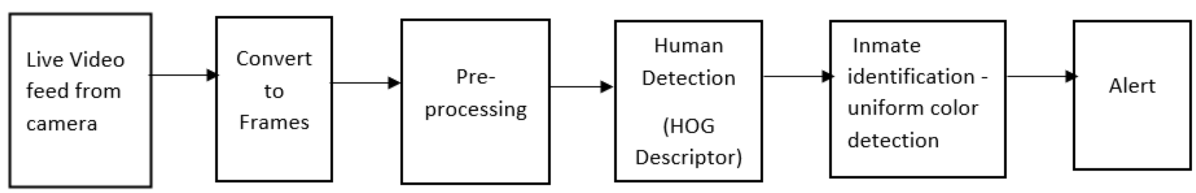


Figure 5.1.1 Block diagram of outdoor camera

5.1.2 Working

The outdoor camera continuously checks whether the inmates are moving out of the institution compound, if detected it generates an alert notifying the caretakers. The basic working idea is to check for the detection of humans in the video frame and then identify the inmates from other people based on the colour of their uniform.

The program first takes each frame of the video and detects humans from the frame. OpenCV features an implementation for a very fast human detection method, called HOG (Histograms of Oriented Gradients). Here we used HOG Descriptor for human detection. It then extracts the human region and converts it from BGR to HSV color-space.Threshold the HSV image for a range of colour required for custom uniform.

HOG is a feature descriptor used in computer vision and image processing for the purpose of human detection. OpenCV has already been implemented in an efficient way to combine the HOG Descriptor algorithm with Support Vector Machine .

The basic idea of the method is the following:

1. The picture is scanned with a detection window of varying size.
2. For each position and size of the detection window, the window is subdivided in cells. The cells are in practice relatively small: they typically contain only a small part of the person to be detected, maybe the side of an arm, or the top of the head.
3. In each cell, a gradient is computed for each pixel, and the gradients are used to fill an histogram: the value is the angle of the gradient, and the weight is the magnitude of the gradient.
4. The histograms of all cells are put together and fed to a machine learning discriminator to decide whether the cells of the current detection window correspond to a person or not.

After capturing the live stream frame by frame and identifying the human region, we are converting each frame in BGR color space to HSV color space. In HSV, it is easier to represent a color than RGB color-space. When RGB is used, it cannot separate color information from luminance but HSV or Hue Saturation Value can do this. This makes it easier when we are working on or need luminance of the image/frame. For color conversion, we use the function cv2.cvtColor(input_image, flag) where flag determines the type of conversion. For BGR to HSV, we use the flag cv2.COLOR_BGR2HSV. In specifying the range , we have specified the range of uniform color. The detectMultiScale method constructs an image pyramid with scale=1.05 and a sliding window step size of (4, 4) pixels in both the x and y direction, respectively. The detectMultiScale function returns a 2-tuple of boxes , or the bounding box (x, y)-coordinates of each person in the image, and weights (the confidence value returned by the SVM for each detection). To solve the problem of multiple overlapping bounding boxes

detected for each person, if one bounding box is fully contained within another, non-maxima suppression is applied to suppress bounding boxes that overlap with a significant threshold.
After applying non-maxima suppression, the finalized bounding boxes coordinates are cropped as an image. To this cropped region, a mask of hsv red layer (taken as the color of uniform here) is applied and bitwise AND operation is done to find if the person is wearing the custom color of uniform. The countNonZero() counts the number of nonzero pixels which is the number of red coloured pixels and a threshold value of 200 is set here. If there are more than 200 pixels of red color, it is assumed that an inmate is present near the compound boundary. Then we are giving an alert by SMS to the caretakers through a platform called Twilio.

5.2 INDOOR CAMERA

5.2.1 Block Diagram

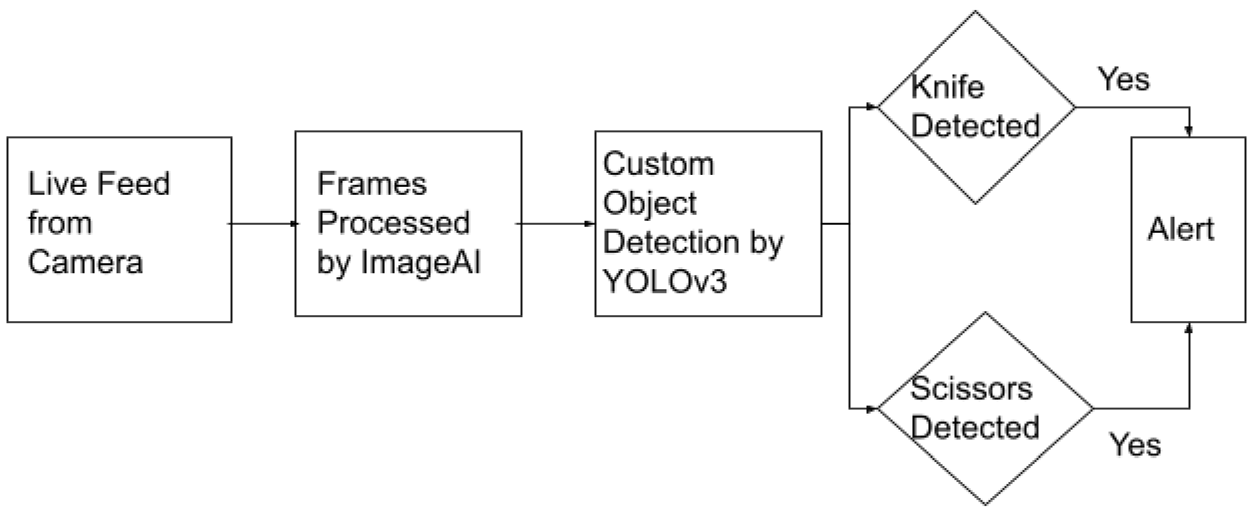


Figure 5.2.1 Block diagram of indoor camera

5.2.2 Working

The indoor camera continuously monitors inmates and the feed is given to ImageAI for further operations. ImageAI provides commercial grade video objects detection features, which include device/IP camera inputs, per frame, per second, per minute and entire video analysis for storing in databases and real-time visualizations for future insights. ImageAI is a python library to build applications and systems with self-contained Deep Learning and Computer Vision capabilities. Dependencies of ImageAI are Python, Tensorflow, OpenCV, Keras.

ImageAI uses the Tensorflow backbone for Computer Vision operations. Tensorflow supports both CPUs and GPUs. All of ImageAI implementations and code can work on any computer system with moderate CPU capacity. However, the speed of processing for operations like image prediction, object detection and others on CPU is slow and not suitable for real-time applications. To perform real-time Computer Vision operations with high performance, GPU enabled technologies are necessary. Hence we used Google Colab platform which gives free access to NVIDIA Tesla K80 GPU that can be used for up to 12 hours continuously.

The VideoObjectDetection class of ImageAI provides a function to detect objects in videos and live-feed from device cameras and IP cameras, using pre-trained models that were trained on the COCO dataset. The models supported are RetinaNet, YOLOv3 and TinyYOLOv3.

Here YOLOv3 is preferred for object detection due to its faster operation with nearly as accurate as Single Shot MultiBox (SSD)[2][13]. It can detect and recognize 80 different kinds of common objects under different classes like people, automobiles, birds & animals, kitchen utensils, household items, foods, furniture, electronics etc. While YOLO struggles with small objects, YOLOv3 has better performance for small objects because of using short cut connections that provides more finer-grained information from the earlier feature map.YOLO v3 makes prediction at three scales, which are precisely given by downsampling the dimensions of the input image by 32, 16 and 8 respectively. The YOLOv3 model file is available on Github and this model file with the .h5 extension is added to Google Drive to be used later.

After that an instance of the VideoObjectDetection is created. Using this instance, different functions can be called to set its properties and to detect objects in a video. The model type is set as YOLOv3 so that object detection tasks are carried out using the pre-trained YOLOv3 model before. After that the path to this model in google drive is set after verification using the setModelPath() function. The model is loaded from the path specified into the object detection instance[10] . The objects we need to detect are knives and scissors only. So the properties of these items are set to True and stored in a variable named custom.

The method detectCustomObjectsFromVideo() performs custom object detection on a video file or video live-feed after the model has been loaded into the instance. The custom variable, along with the path to the input video file or camera (using OpenCV's VideoCapture() function), path to which the detected video will be saved (in avi format) etc are given as parameters to this function. Further, frames per second can be set to desired fps for the detected video.The default value is 20.

Other controllable but optional parameters include *log_progress* that shows the progress of the video or live-feed as it is detected in the CLI, *minimum_percentage_probability* to determine the integrity of the detection results, *display_percentage_probability* to show/hide the percentage probability of each object detected in the detected video, *display_object_name* to show/hide the name of each object detected in the detected video, *save_detected_video* to save the detected video etc.

Lowering the value of minimum_percentage_probablilty shows more objects while increasing the value ensures objects with the highest accuracy are detected. The default value is 50.

ImageAI also provides detection speeds for all object detection tasks to reduce the time of detection at a rate between 20% - 80% having just slight changes but accurate detection results. The available detection speeds are normal(default), fast, faster , fastest and flash.

Separate functions can be specified to be repeated every second, every minute, every frame or to the whole video. The per_minute_function allows parsing in the name of a function that is defined separately. Then for every minute of the video that is detected, the function parsed into the parameter will be executed and analytical data of the video will be parsed into the function. The data returned can be visualized or saved in a NoSQL database for future processing and visualization. Similarly there is a per_second_function and per_frame_function.

Here YOLOv3 is constantly searching for the presence of knives or scissors in video. The method passed to per_minute_function is sendAlert method, set in such a way that in every minute, confidence of the prediction is taken as a threshold value and if it is greater than 60%, the program will set the alert flag and detected flag and an alert is sent to the caretaker via sms using twilio platform. It then waits for 2 minutes before detecting again.

CHAPTER 6

FUTURE SCOPE

1. Incorporating fire detection in indoor cameras so that necessary precautions can be made before widespread damage. Fire accidents are the most commonly occurring disasters nowadays. Fire, especially in buildings, can spread quickly and cause great loss of life and property. Due to the rapid increase in the number of fire accidents, the fire alarm systems form an integral part of the necessary accessories in any sort of construction. Therefore, early fire detection and warning is imperative. In order to mitigate the number of fire accidents, a large number of methods have been proposed for early fire detection to reduce the damage caused by such accidents. Apart from the problem of early fire detection, present fire alarm systems also prove to be inefficient in terms of the false triggering of the alarm systems.
2. Including fight or violence detection so that physical harm can be reduced. Autism patients have an unpredictable nature and tend to get aggressive sometimes and may even attack others. Such violent behaviour needs to be monitored and their fight should be broken before anything fatal happens to anyone.Thus recognizing fights and aggressive behavior[7] in video is an increasingly important application area & such capability is extremely useful in other video surveillance scenarios like in prisons, psychiatric or elderly centers too. Action recognition techniques that have focused largely on individual actors and simple events can be extended to this specific application.
3. Emotion recognition[7] - Children with autism spectrum disorder have difficulty in understanding the emotional and mental states from the facial expressions of the people they interact with. The inability to understand other people's emotions will hinder their interpersonal communication. Though many facial emotion recognition algorithms have been proposed in the literature, they are mainly intended for processing by a personal computer, which limits their usability in on-the-move applications where portability is desired. The portability of the system will ensure ease of use and real-time emotion recognition and that will aid for immediate feedback while communicating with caretakers.

CHAPTER 7

OUTPUT

7.1 Detection of Weapons

Weapon detection or primarily, knife and scissors detection in real-time[5] is done using a pre-trained YOLOv3 model that is customized by ImageAI to detect only these 2 weapons.

Figure 7.1.1 Detection of knife

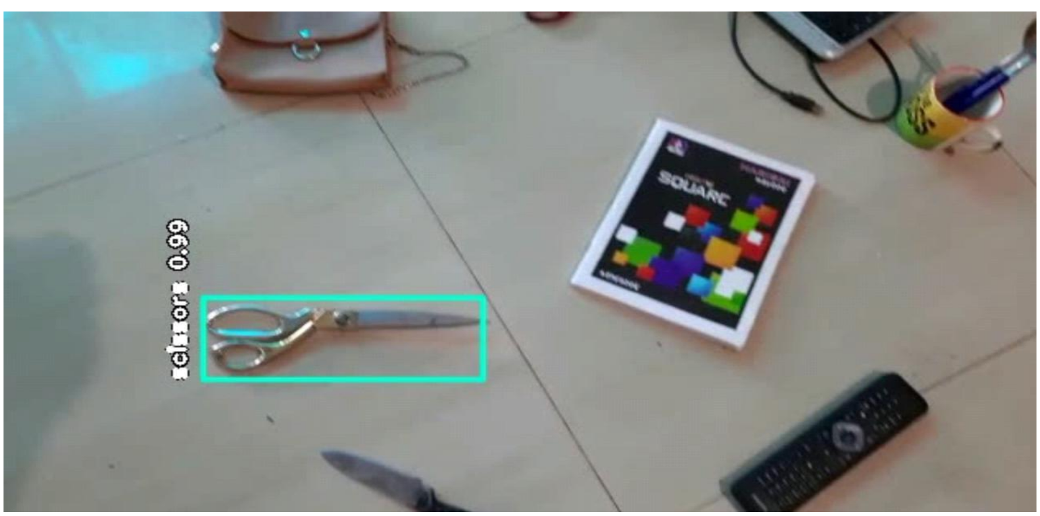


Figure 7.1.2 Detection of scissors

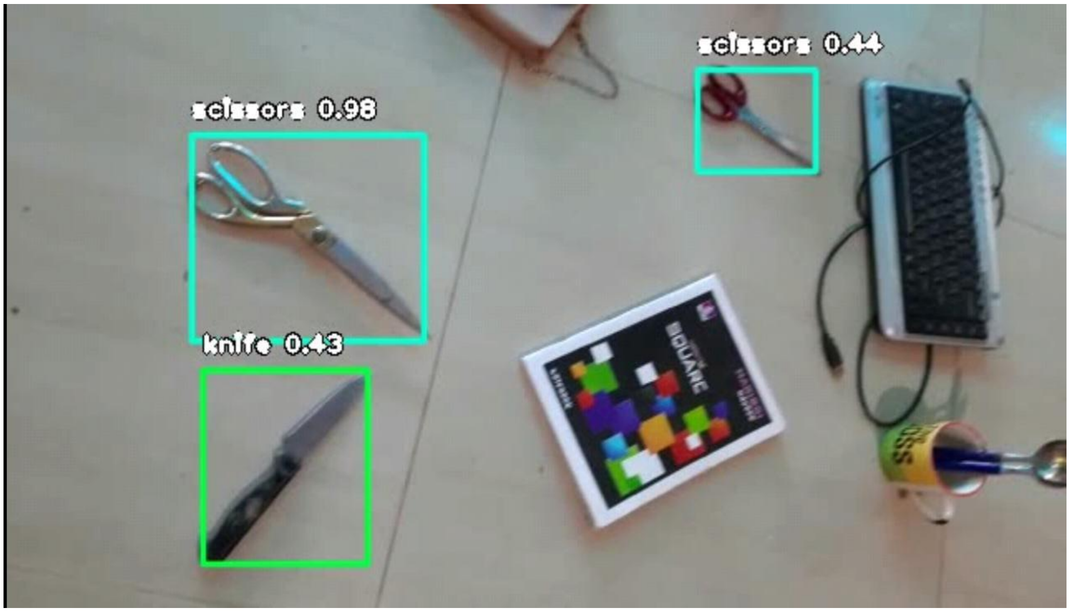


Figure 7.1.3 Detection of weapons

7.2 Inmate Detection

Using a HOG descriptor, a person is detected from the video frame and using color detection[9], the inmate is identified based on their uniform color .

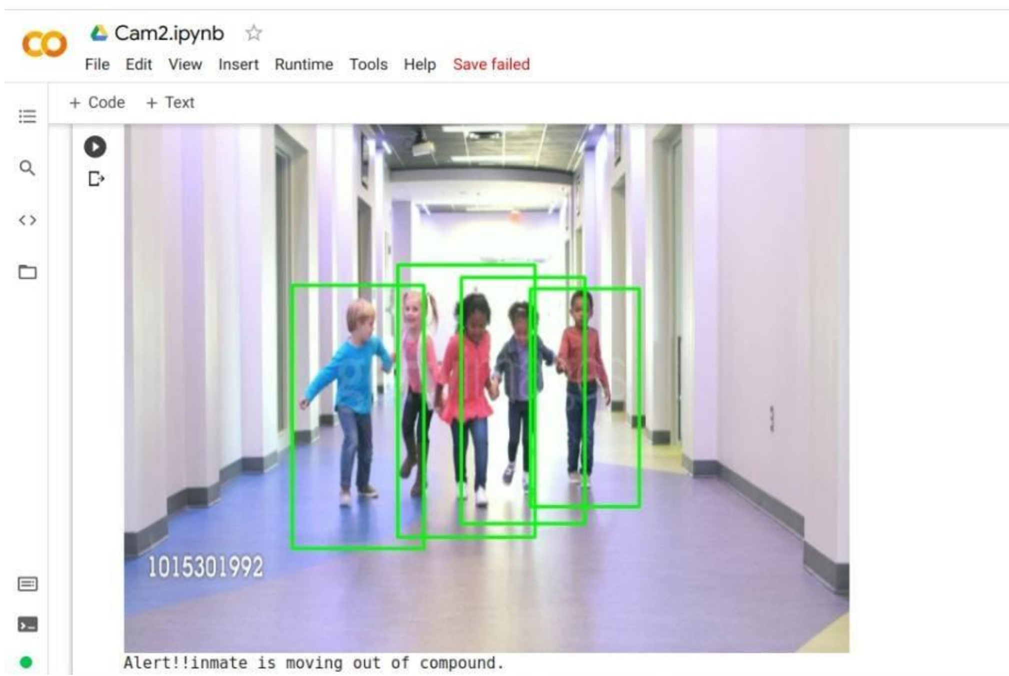


Figure 7.2.1 Inmate detection 1 - crossing boundary

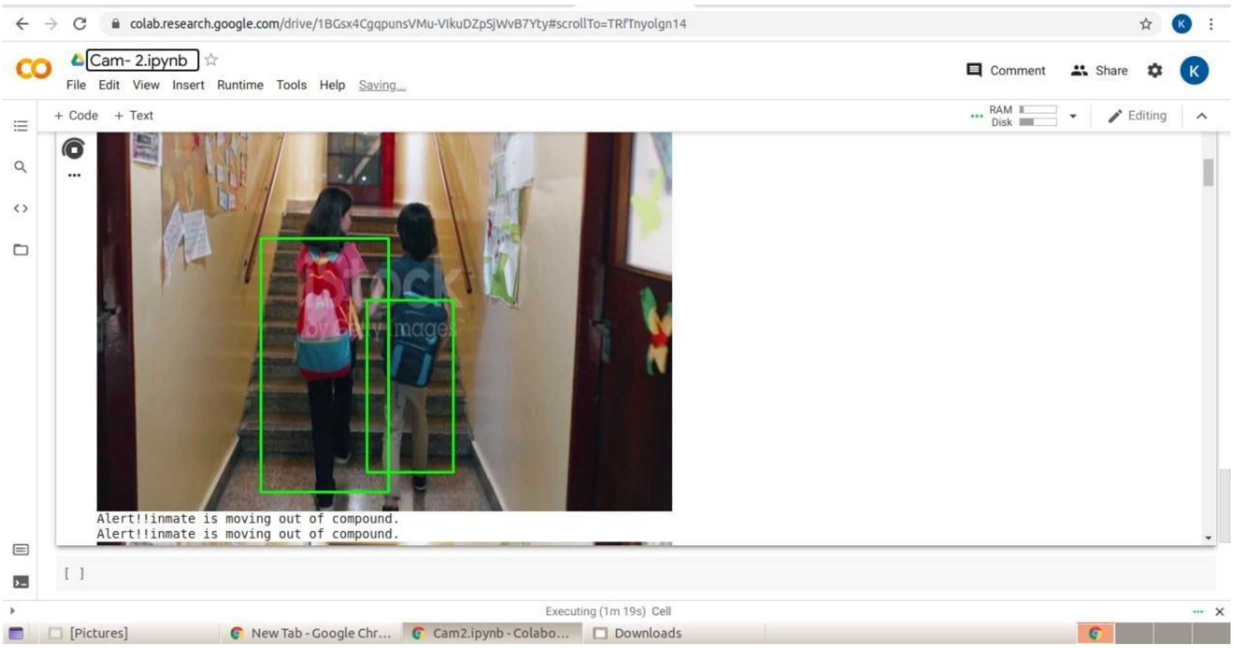


Figure 7.2.2 Inmate detection 2 - crossing boundary

7.3 Alert Message

Alerting caretakers via sms when inmates possess harmful objects or when they try to move out of the compound is accomplished with the help of Twilio platform.

Figure 7.3.1 Twilio message - Weapon detected

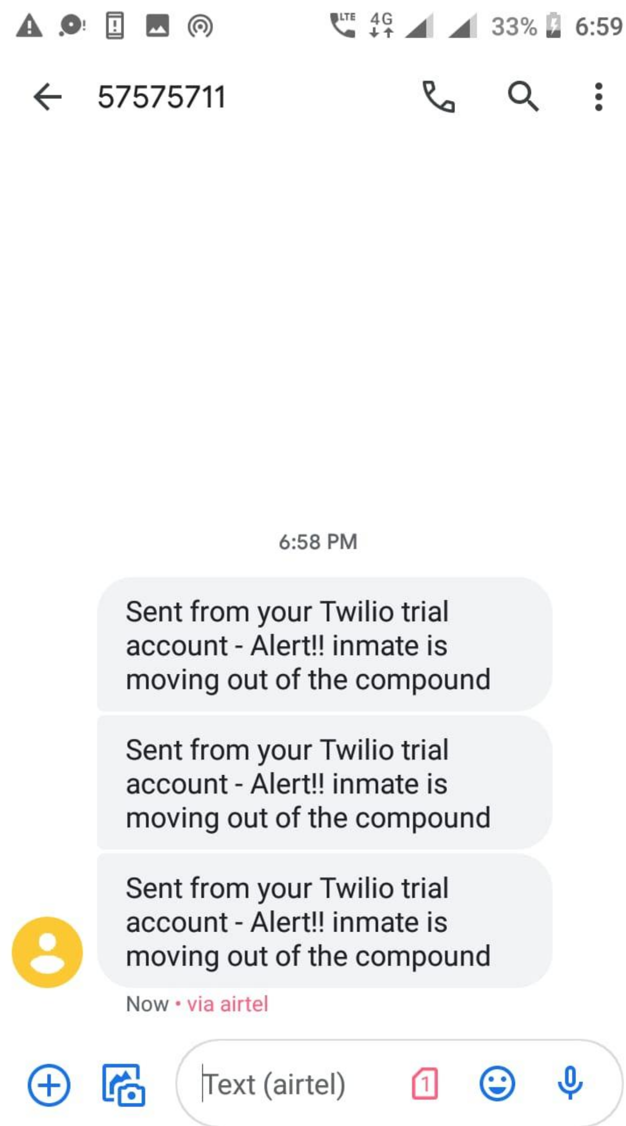


Figure 7.3.2 Alert message- Inmate moving out

CHAPTER 8

CONCLUSION

The health of the soul and the body is one of the important issues requiring more attention. Autism is one of the most mysterious and complex disorders that make social communications difficult.Autism Spectrum Disorder (ASD)s[12] continue to be an important public health concern.

Considering the wide spread of technology in various fields of information, the use of methods such as image processing can be one of the effective methods in the field of treatment, especially neurological and psychiatric disorders.Image processing technique plays an important role in the detection.

In this project we have proposed a surveillance system in autism centres using image processing. It enables the caretakers of ASD centres to monitor the inmates in a much effective manner in a virtual reality environment in spite of their potent impairments.

APPENDIX

Outdoor Cam Code

```
import numpy as np
import cv2
from google.colab.patches import cv2_imshow
from imutils.object_detection import non_max_suppression
import numpy as np
import imutils
import cv2
import time
from twilio.rest import Client


#assigning id and token from twilio platform
sid = "ACaf6a5dc25975e7ab90d8566fe17ac4a8"
token = "a510c53a7c502720060f7b70b333382b"
client = Client(sid,token)


from google.colab import drive
drive.mount('/content/gdrive')
alert=False
# initialize the HOG descriptor/person detector
hog = cv2.HOGDescriptor()
hog.setSVMDetector(cv2.HOGDescriptor_getDefaultPeopleDetector())

# open video stream
cap = cv2.VideoCapture('/content/gdrive/MyDrive/Colab Notebooks/uniformdet.mp4')

# the output will be written to uniform-detector.avi
```

```
out = cv2.VideoWriter(
    '/content/gdrive/MyDrive/Colab Notebooks/uniform-detector',
    cv2.VideoWriter_fourcc(*'MJPG'),15.,(640, 480))

while (True):
    # Capture frame-by-frame
    ret, frame = cap.read()
    # resizing for faster detection
    frame = cv2.resize(frame, (640, 480))
    # using a grayscale picture for faster detection
    gray = cv2.cvtColor(frame, cv2.COLOR_RGB2GRAY)

    # detect people in the image
    # returns the bounding boxes for the detected objects
    boxes, weights = hog.detectMultiScale(frame, winStride=(8, 8))

    boxes = np.array([[x, y, x + w, y + h] for (x, y, w, h) in boxes])
    # to reduce overlapping bounding boxes
    persons = non_max_suppression(boxes, probs=None, overlapThresh=0.65)

    for (xA, yA, xB, yB) in persons:
        #crop the part where person is detected
        cropped = frame[yA:yB, xA:xB]

        #Blur the image
        blur = cv2.GaussianBlur(cropped, (21, 21), 0)
        #RGB to HSV Conversion
        hsv = cv2.cvtColor(blur, cv2.COLOR_BGR2HSV)
        #Lower and Upper HSV values of Red
        lower1= [0,70,50]
        upper1 = [11,255,255]
```

```
        lower1= np.array(lower1, dtype="uint8")
        upper1= np.array(upper1, dtype="uint8")
        #Sets the Color Range and returns a binary mask
        mask = cv2.inRange(hsv, lower1, upper1)
        #AND operation of mask and frame to detect red color
        output = cv2.bitwise_and(cropped, hsv, mask=mask)

        #to find the number of nonzero pixels in the array matrix/ the red pixels
        pers= cv2.countNonZero(mask)
        # threshold value
        if int(pers) > 200:
          print("Alert!! inmate is moving out of the compound")
          #sending alert SMS when inmate comes in the field of view of camera
          to="+919995943886",
          from_="+15709895266",
           body= "Alert!! inmate is moving out of the compound")
   print(message.sid)

   # display the detected boxes in the colour picture
        cv2.rectangle(frame, (xA, yA), (xB, yB),
                      (0, 255, 0), 2)
  # Write the output video
 out.write(frame.astype('uint8'))

# Display the resulting frame
 cv2_imshow(frame)
 time.sleep(120)



if cv2.waitKey(1) & 0xFF == ord('q'):
        break
```

```
cap.release()
out.release()
# close the window
cv2.destroyAllWindows()
cv2.waitKey(1)
```

Indoor Cam Code

```
#installing libraries
!pip3 install https: // github.com / OlafenwaMoses / ImageAI / releases / download / 2.0.2 /
imageai - 2.0.2 - py3 - none - any.whl
!pip install -q opencv-python
!pip install -q pillow

#importing libraries
from imageai.Detection import VideoObjectDetection
import matplotlib as plt
import tensorflow as tf
import numpy as np
import scipy
import keras
import h5py
import torch
import time
from tensorflow.python.framework.ops import disable_eager_execution
from PIL import Image
from twilio.rest import Client
```

```
#assigning id and token from twilio platform
sid = "ACaf6a5dc25975e7ab90d8566fe17ac4a8"
token = "a510c53a7c502720060f7b70b333382b"
client = Client(sid,token)

#creating an instance of Video object detection
detector = VideoObjectDetection()

#setting model type of the object detection instance to YOLOv3 model
detector.setModelTypeAsYOLOv3()

#mount the drive to import yolo.h5 and video
from google.colab import drive
drive.mount('/content/gdrive')

#function to accept a string which must be the path to the model file downloaded
detector.setModelPath("/content/gdrive/MyDrive/Colab Notebooks/yolo.h5")
disable_eager_execution()

#function to load model from the path specified in the function call to our object detection instance
detector.loadModel()
custom = detector.CustomObjects(knife=True, scissors=True)

 alert=False
def sendAlert(frame,output_array,count):
  detected=0

 while(!alert):
   #output_array is a list of dictionaries
```

```
 #Dictionary has boxpoints, name and percentage probability as keys

  for dt in output_array:
        for key in dt:
          if key==percentage_probability:
              if dt[key]>60:
                    detected +=1
                    if detected>5 and alert==False:
                           alert=True
                  #sending alert SMS when weapon is detected
                           message = client.messages.create(
                                     to="+919995943886",
                                     From_="+15709895266",
                                     body= "Weapon is detected")
                           print(message.sid)
                           time.sleep(120)
                           alert=False

video_path = detector.detectCustomObjectsFromVideo(custom_objects=custom,
                                 input_file_path="/content/gdrive/MyDrive/Colab Notebooks/yolodet.mp4",
                                 output_file_path="/content/gdrive/MyDrive/Colab Notebooks/yoloh-detected8.mp4",
                                 frames_per_second=12, minimum_percentage_probability=20,
                                 log_progress=True, display_percentage_probability=True,
                                 per_minute_function=sendAlert, return_detected_frame=True)
```

REFERENCES

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

[2] J. Li, Q. Hou, J. Xing and J. Ju, "SSD Object Detection Model Based on Multi-Frequency Feature Theory," in *IEEE Access, vol. 8, pp.* 82294-82305, 2020, doi: 10.1109/ACCESS.2020.2990477.

[3] G. Chandan, A. Jain, H. Jain and Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1305-1308, doi: 10.1109/ICIRCA.2018.8597266.

[4] Stolovy, Tali & Melamed, Yuval & Afek, Arnon. (2015)." Video Surveillance in Mental Health Facilities: Is it Ethical?." *The Israel Medical Association journal* : IMAJ. 17. 274-6.

[5] S. Xu and K. Hung, "Development of an AI-based System for Automatic Detection and Recognition of Weapons in Surveillance Videos," 2020 *IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE),* Malaysia, 2020, pp. 48-52, doi: 10.1109/ISCAIE47305.2020.9108816.

[6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[7] Doras Oladayo Esan, Pius A. Owolawi, Chunling Tu. Detection of Anomalous Behavioural Patterns In University Environment Using CNN-LSTM. In *IEEE 23rd*

International Conference on Information Fusion, FUSION 2020, Rustenburg, South Africa, July 6-9, 2020. pages 1-8, IEEE, 2020. [doi]

[8] P.Priyanka, T.Sivapriya, Sd.Laheen, V.Lakshmi Lahari, " SMS Notifications using Raspberry pi "ISSN: 2350-0328 *International Journal of Advanced Research in Science, Engineering and Technology* Vol. 6, Issue 3, March 2019 Copyright to IJASET, www.ijarset.com 8515

[9] A. S. Silva, F. M. Q. Severgnini, M. L. Oliveira, V. M. S. Mendes and Z. M. A. Peixoto, "Object Tracking by Color and Active Contour Models Segmentation," in *IEEE Latin America Transactions,* vol. 14, no. 3, pp. 1488-1493, March 2016, doi: 10.1109/TLA.2016.7459639.

[10] Z. Chen, R. Khemmar, B. Decoux, A. Atahouet and J. Ertaud, "Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility," 2019 *Eighth International Conference on Emerging Security Technologies (EST), Colchester, United Kingdom,* 2019, pp. 1-6, doi: 10.1109/EST.2019.8806222.

[11] Liu W. et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham.

[12] A. Chorianopoulou, E. Tzinis, E. Iosif, A. Papoulidi, C. Papailiou and A. Potamianos, "Engagement detection for children with Autism Spectrum Disorder," 2017 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 5055-5059, doi: 10.1109/ICASSP.2017.7953119.

[13] N. I. Hassan, N. M. Tahir, F. H. K. Zaman and H. Hashim, "People Detection System Using YOLOv3 Algorithm," 2020 *10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE),* 2020, pp. 131-136, doi: 10.1109/ICCSCE50387.2020.9204925.