# Real-time Smart Navigation System for Visually Impaired People

Project ID: 2022-256

Project Proposal Report

Madushan W. A.   IT19042152

Supervisor: Mrs. Sanjeewi Chandrasiri

B.Sc(Hons) Degree in Information Technology Specializing in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2022

# Real-time Smart Navigation System for Visually Impaired People

Project ID: 2022-256

Project Proposal Report

B.Sc(Hons) Degree in Information Technology Specializing in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2022

# DECLARATION

“I declare that this is our own work, and this proposal does not incorporate without acknowledgement  any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).”

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

| Name | Student ID | Signature |
|---|---|---|
| Madushan W. A. | IT19042152 | |

Signature of the supervisor:

## ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude for the immense support, guidance and motivation provided by my supervisor Mrs. Sanjeewi Chandrasiri which always helped me for the successful completion of my undergraduate research. Her enthusiasm towards research motivated us to engage in competitions and to meet new people from the industry which otherwise would not have been possible.

Moreover, our special thanks go to our seniors for the greatest guidance and support given us in order to end this research successfully. Furthermore, we would like to thank our parents for their patience, time and providing resources to acquire the needful and for other expenses. Finally, I express my sense of gratitude to my teammates, family, friends, to one and all, who directly or indirectly have extended their support throughout this project.

# ABSTRACT

The main focus of this project is to give a better and more efficient solution for navigation issues of the visually impaired people. Visual impairment has always made it difficult for someone to go about their daily tasks normally. Their problems are countless, especially in public areas. It has always been difficult for people to read traffic signs to protect their safety when traveling on roads and crossings. The Vision Atlas of the International Agency for the Prevention of Blindness estimates that there are 295 million people with moderate-to-severe visual impairment and 43 million people who are completely blind[1]. For a long lime, white cane and guide dogs have become the solution for this matter[2]. But they can give a limited assistance for finding the way to a remote location only. The objective is to develop a portable, self-contained device that will let visually impaired people move through both familiar and unfamiliar places on their own without the use of guides. Several electronic devices are currently available for providing guidance to a remote location, but these tend to be expensive, or can be used for only one purpose. This introducing system is capable of identifying obstacles and potholes, walk lanes, faces of known individuals and road signs. Also, it contains an alternative solution for the accessible pedestrian signals. It allows user to make the decision whether it is safe to cross the road, or they should wait for the next turn. Furthermore, the system informs the blind person via voice through the headphone. Through this system most of the problems and challenges of visually impaired people faced during their navigation have solved. System allows them to travel with increased independence, safety, and confidence.



**Keywords**

Visually impaired, Blind Navigation, Image processing

# TABLE OF CONTENT

# LIST OF FIGURES

# LIST OF ABBREVIATION

| | |
|---|---|
| CNN | Convolutional Neural Network |
| IOT | Internet Of Things |
| ICT | Information and Communication Technology |
| AI | Artificial Intelligent |
| ANN | Artificial Neural Network |
| CPU | Central Processing Unite |
| | |
| | |

# 1. INTRODUCTION

## 1.1 BACKGROUND LITERETURE

Even a straightforward travel task involves a lengthy list of related tasks. Mobility and environmental access are the two main categories of the subtasks in the trip activity[3]. While environment access consists of danger reduction and information/sign, mobility itself can be separated into obstacle avoidance and orientation/navigation. The majority of the travel-related subtasks are based on visual data. People who are sighted mostly rely on their sense of sight for this[4]. People who are visually challenged can only use their sense of sight to a limited or perhaps nonexistent extent. Therefore, in order to complete various travel activity subtasks, people with visual impairments need assistance from assistive technology. We concentrate on creating helpful technology for visually impaired persons to avoid obstacles because it has long been thought of as a crucial component of assisted mobility. Obstacle detection and obstacle warning are two problems that obstacle avoidance technology needs to solve. The former refers to the process of communicating barrier information to those who are visually blind, whereas the latter refers to the perception of potentially dangerous things in the environment in advance.

Independence is crucial for accomplishing life's aspirations, objectives, and ambitions. Individuals with visual impairments find it challenging to travel on their own. Around 285 million individuals have visual impairments, of which 246 million have a decline in visual acuity and 39 million are blind, according to the World Health Organization. Nearly 90% of visually handicapped people reside in developing nations. Certain abilities connected to visual function are severely affected by these visual impairments:

- The daily tasks (which demand vision at a medium distance)
- Communication, reading, and writing (which call on close-up and normal distance vision)
- Evaluation of space and displacement (which calls for a long-distance vision)
- The pursuit of a task that requires sustained maintenance of vision.

There are 43 million people who are completely blind, according to records. Additionally, there are 295 million people who have moderate to severe visual impairment. According to the World Health Organization (WHO) Global Action Plan for Universal Eye Health, population-based data on visual impairment are necessary to estimate the need for services, assess service delivery, and highlight priority that need to be addressed[4]. A nationwide poll was started since Sri Lanka lacked any national statistics. It covers five main conditions, namely cataract, glaucoma, blindness in children, diabetic retinopathy, refractive errors, and low vision. On May 6 and 7, 2016, VISION 2020 Sri Lanka and the Ministry of Health hosted a nationwide workshop on an action plan for eye health in cooperation with IAPB Southeast Asia[5]. The government offers free medical care, which includes vision care. The major goal of the workshop was to develop modification ideas for Sri Lanka's National Eye Health Action Plan 2013–2018 based on the results of their most current blindness survey, which was conducted in 2014–2015, and to include monitoring indicators for the WHO Global Action Plan in it. According to the report, 1.7% of people in Sri Lanka aged 40 and older are blind. Similar to this, the survey found that among the study population, there were 1.6% and 15.4% cases of severe vision impairment and visual impairment. In Uva province

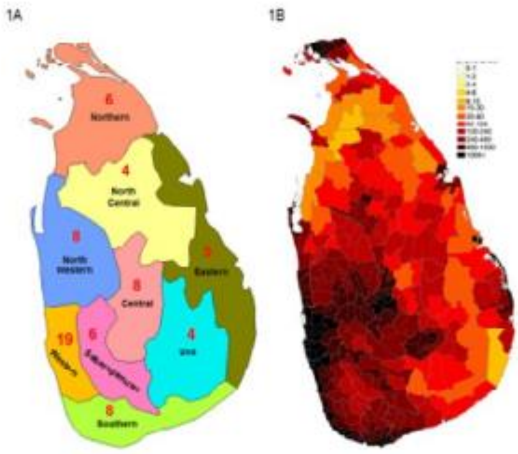


*Figure 1: survey clusters in each Province.*

(Southeast), the prevalence of blindness is as high as 2.9%, whereas it is as low as 0.29% in the South. Nearly 67% of all blindness is caused by cataracts, which is the major cause of blindness. With a visual acuity cut off of 3/60, it was discovered that 85.4% of people have undergone cataract surgery. 59.7% and 75.1%, respectively, of the eyes that underwent cataract surgery had good presenting and best corrected visual acuity results. The ability of those who are blind to engage with their surroundings will be considerably increased by the creation of a successful visual information system.

**Sampling**

The survey selected a group of individuals aged 18 years to estimate the prevalence of disability and a nationally representative sample of adults aged 40 years to estimate the prevalence of blindness. There were 25 districts and all nine provinces covered. The districts served as the main sample units.

**Sample size**

Based on earlier data from the South Asia region, the sample size was calculated using the following parameters: prevalence of blindness (presenting vision) among people aged 40 years: 2.5%; confidence interval: 95%; acceptable error: 0.02; precision: 80%; design effect: 1.5; and response rate: 85%.

There were around 6,600 samples. This was not large enough to produce correct estimates at the district level, but it was intended to yield accurate estimates for the national magnitude of blindness and visual impairment, including the primary causes of blindness, ocular morbidity, and disability among persons aged 40 years.

| Province/District | Number of clusters | | |
|---|---|---|---|
| | Total | Rural | Urban |
| Western Province | 19 | 12 | 7 |
| Colombo | 8 | 2 | 6 |
| Gampaha | 7 | 6 | 1 |
| Kalutara | 4 | 4 | 0 |
| Central Province | 8 | 7 | 1 |
| Kandy | 4 | 3 | 1 |
| Matale | 2 | 2 | 0 |
| NuwaraEliya | 2 | 2 | 0 |
| Southern Province | 8 | 8 | 0 |
| Galle | 3 | 3 | 0 |
| Matara | 3 | 3 | 0 |
| Hambantota | 2 | 2 | 0 |
| Northern Province | 6 | 6 | 0 |
| Jaffna | 2 | 2 | 0 |
| Kilinochchi | 1 | 1 | 0 |
| Mannar | 1 | 1 | 0 |
| Vavuniya | 1 | 1 | 0 |
| Mullaitivu | 1 | 1 | 0 |
| Eastern Province | 5 | 5 | 0 |
| Batticaloa | 2 | 2 | 0 |
| Ampara | 2 | 2 | 0 |
| Trincomalee | 1 | 1 | 0 |
| North Western Province | 8 | 8 | 0 |
| Kurunegala | 5 | 5 | 0 |
| Puttalama | 3 | 3 | 0 |
| North Central Province | 4 | 4 | 0 |
| Anuradhapura | 3 | 3 | 0 |
| Polonnaruwa | 1 | 1 | 0 |
| Uva Province | 4 | 4 | 0 |
| Badulla | 3 | 3 | 0 |
| Monaragala | 1 | 1 | 0 |
| Sabaragamuwa Province | 6 | 6 | 0 |
| Ratnapura | 3 | 3 | 0 |
| Kegalle | 3 | 3 | 0 |
| Total | 68 | 60 | 8 |

The white cane was the first assistance device for avoiding obstacles. However, it is not typically utilized to find barriers higher than knee height. A variety of obstacle avoidance systems are now readily available for visually impaired people thanks to recent advancements in sensor technology[5]. The majority of studies, however, concentrate on obstacle detection; obstacle warning is not thoroughly researched.

**Mobility challenges**

Different obstacle detection and warning systems can be categorized according to how the obstacles are discovered and how the user is made aware of them[6]. Many visually impaired people can navigate independently to substantial degrees with the help of a long white cane. In a poll conducted by the National Federation of the Blind, 280 participants reported that 94% of them use a long white cane. Of these, 25% say they have been hurt indoors, while 61% say they have been hurt outside[7].

*Figure 2: white cane*

The long white cane is a very useful mobility aid for the blind despite being a rather straightforward gadget. The cane offers data of the user's journey in two different ways when used appropriately. The user is reassured that the path is free in front by the long white cane. Second, it shields the user from slipping down a curb or stumbling over an obstruction. The white cane is selected as one of the bases for the prototype system presented in this thesis due to its capacity to give valuable data and its widespread acceptance among the blind people. In the unlikely case this system cannot identify an object, the long white cane adds an extra layer of security to help people avoid potentially hazardous situations.

Various technologies, including Wi-Fi, RFID, laser, ultrasound, and cameras, have been utilized to help blind individuals avoid environmental hazards in the literature. Only vision-based techniques that are reasonably similar to this work in this paper are presented in this part. Depending on how the obstacles are found and how the user is informed of them, different obstacle detection and warning techniques can be divided into different categories.

## 1.2 RESEARCH GAP

The following systems were discovered during the search for comparable methodologies used to address the problem prior to moving forward with the project. There was no system in Sri Lanka that dealt with the aforementioned problems directly, although there were several systems that dealt with the major issue, "Navigation Systems for People with Visual Impairments."

At least a few attempts have been made to help those who are vision impaired with their daily routines. Most initiatives improve the tools that blind people already use with new cutting-edge technology. This keeps the amount of stuff blind people need to carry the same. There are also remote human assistance services, in which blind people can receive telecommunications-based aid for a variety of tasks. The latter is typically used by big organizations.

A text-to-speech application was developed in 2010 by the Hewlett-Packard and Mobile Speak Pocket businesses to make mobile phone technologies accessible to people with visual impairments [8]. This software converts the visual data from the phone into speech. Users may browse the internet, chat with friends, and move around the many screens of their mobile phones when equipped with the phone's functioning buttons, all without the need for perception. The Intel Reader [9] devices, another effort from Intel in a related direction, have text recognition software that reads material aloud. When combined with object recognition systems, this is quite beneficial. Because text alone can occasionally be used to display navigational information. The Intel Reader takes text-containing images and converts them into audio by using text recognition and image processing techniques.

By supporting people who are blind with grocery shopping, Foo et al. presented yet another intriguing piece[10]. It can be highly expensive for grocery businesses to help persons who are blind buy food. They may be able to buy food on their own with the use of this application. Given that eating is a daily necessity, this might make life easier for persons who are blind or visually impaired. The object identification, audio localization, and text-to-speech notifications employed in this program are fundamental computer vision techniques. One of the hallmarks of the Grozi project was its ability to help people find aisles using sound localization and identify items and aisle segments using object recognition. A multi-dimensional microphone system was utilized to pinpoint the location of the actual 3-dimensional scene in each isle after a specific audio for that isle was repeatedly played. Although this is a workable method, it can only be used in quiet settings. Which is highly unlikely given how noisy grocery stores frequently are during business hours. It may be more efficient to use an image-based recognition technology that can read QR codes put at the start of aisles.

The authors of [11] suggested a blind stick for use in navigation by people with vision impairments. They created the algorithm and integrated it into the smart pole for accurate path finding and obstacle identification. The buzzer is utilized as an alert, while the ultrasonic sensor is used to identify obstacles. The method can only be used for a short distance, and a visually impaired person can only locate inside that short distance. Additionally, the stick gives those with disabilities their own sense of independence.

In [12], they created a low-cost, approachable smart blind navigation system that is used for the mobility of visually impaired persons over short distances. An infrared sensor device is used to calculate the barrier's reflected range. The microcontroller manages the system as a whole and the reflected range. The development of an efficient smart stick for blind persons [13] uses an infrared sensor to track stair movement up and down. The ultrasonic sensing element is used to identify obstacles, and the microprocessor controls the entire system. The entire integrated system is wired, and wireless functionality may be added in the future.

Ultrasonic sensors were used in the development of the Ultrasonic Blind Walking Stick [14]. This instrument's primary goal is to eliminate the drawbacks of traditional sticks. The obstruction that is present in a visually impaired person's line of sight is detected by the ultrasonic sensor component. This device has an accurate understanding of uphill and downhill terrain. This blind stick can spot obstructions and obstacles, especially when people with disabilities are travelling alone in more muddy areas. Mishra and Koley [15] created a relatively affordable tool to locate and track down a person who is blind or visually impaired. The goal is to provide low-cost devices and fully reduce the requirement of guides for those with vision impairments in rural areas who need to find their way around. The ultrasonic sensor is utilized to detect obstacles, while the GPS tracking device provides location data. For data transmission and receiving, the system also included voice-based system communication protocols.

| Feature | Research A [1] | Research B [2] | Research C [3] | Research D [4] | Research E [5] | Smart Navigation Application |
|---|---|---|---|---|---|---|
| Object detection | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Close object warning | ✗ | ✗ | ✓ | ✗ | ✗ | ✓ |
| Voice assistant | ✗ | ✓ | ✗ | ✓ | ✓ | ✓ |
| Surface detection | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Pothole detection | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ |

*Figure 3: Research gap*

Most of the research and applications have mentioned earlier is not used to detect surface for blind people. Most of the research have done to develop new machine learning models and identify new technologies which can detect basic objects and obstacles more accurately. Research E [16] is the only one which can detect objects and pothole, and which is able to give the voice assistant too. Also, research C [17] is the only one which can identify close objects and notify the user about it. Any of the existing solutions are not able to detect surface of the object and notify user about it. In the proposed system, surface of the path or an obstacle will be captured by the camera and analyze them using image processing and machine learning technologies. Then the user will be notified about dangerous and slippery surfaces.

# 2. RESEARCH PROBLEM

By January 2022[18], there will be 7.9 billion people living in the world, and 2.2 billion of them will have some form of vision impairment, such as close vision impairment, distance vision impairment, partial blindness, or complete blindness [19]. Approximately 45 million people worldwide are considered to be blind, and 246 million more have limited eyesight [20]. Additionally, 82% of all visually impaired individuals are over the age of 50 [21], and 87% of them reside in underdeveloped nations.

According to the figures above, a sizeable portion of the population is affected by blindness or other types of vision problems. Additionally, it is evident that the bulk of them are over 50, which results in a disability brought on by their visual restrictions. One of the main causes of impairment, particularly in older persons, is vision loss [22]. This condition will lower quality of life, which will raise melancholy, anxiety, and other illnesses owing to restricted motion. According to Pawel Strumillo, vision loss is the most severe sensory impairment, accounting for about 5% of all cases. 90% of an individual's capacity for multisensory perception has been lost [21].

*Figure 4: Asus Xtion Pro Live is attached to the white cane handle*

However, due to the fact that the majority of visually impaired people live in developing nations and struggle to make ends meet, these traditional methods of mobility aids for the blind can only be as discussed above. As a result, it makes it more difficult for them to carry out their daily tasks and restricts their ability to integrate into society. A visually impaired individual faces difficulties with even the most fundamental tasks, including autonomously moving from one location to another, which further limits their ability to participate in society and the economy [25]. Mobility assisting devices, such as braille signs and labels, magnifiers, special lights to aid low vision, and canes stretched so far, though still far from perfect, are used to treat this issue and attempt to ease the sensory handicap. Modern technology has advanced with numerous goods to close this gap. Two significant study fields were formed within this framework. To enable independent, secure movement from one place to another, the first one deals with the recognition of surfaces and potholes, while the second one deals with the sensing things around the user[24][23]. Since the currently available technical solutions for the above-mentioned objectives are expensive, aim of this innovation is to develop a cheaper alternative for this[24].

In earliest stages of this research, an informal survey of a visually impaired focus group was conducted. The focus group identified various challenges associated with blind navigation. The obstacles faced by the blind and visually impaired are numerous and complex, even with the aid of a conventional cane. Many of the blind people we spoke with still have some vision. One gentleman spoke about how retinitis pigmentosa caused him to lose his vision. He described his eyesight as having a severe loss of peripheral vision and having a lot of trouble with Night vision, contrast, and color. Any of the people is unable to read plain text or efficiently utilize a cell phone and computer without assistive technology. Thankfully, conference attendees acknowledged use a variety of assistive technology to access printed content. Reading a computer or cell phone screen is achievable by inverting the contrast and enlarging the text.
To access printed material, people without light perception employ optical character recognition (OCR) and other voice-over technology. The majority of those attending the meeting used a long white cane for navigating. The long white cane has a number of advantages, according to attendees. Navigation greatly benefits from its primary purpose of detecting the position, size, and kind of impediments. By marking curbs and steps, the cane creates a clear path and helps prevent falls. It is used to locate door openings, which can be challenging if there are surrounding large windows or if the door is constructed of glass. Attendees did mention, however, that they frequently found the cane to be insufficient for all of their navigational needs.

*Figure 5: RGB image*

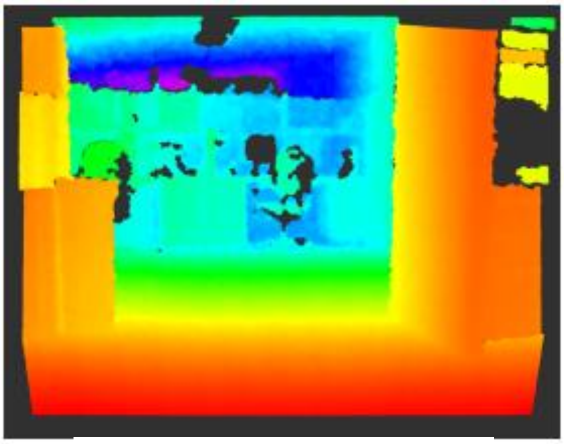


*Figure 6: Point cloud representation in 1*

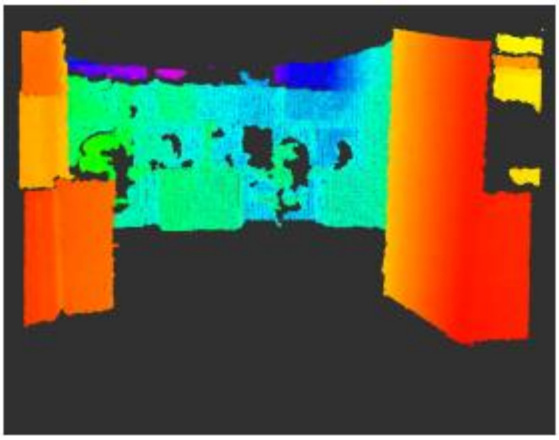


*Figure 7: Point cloud in 2 with the ground and above head points removed*

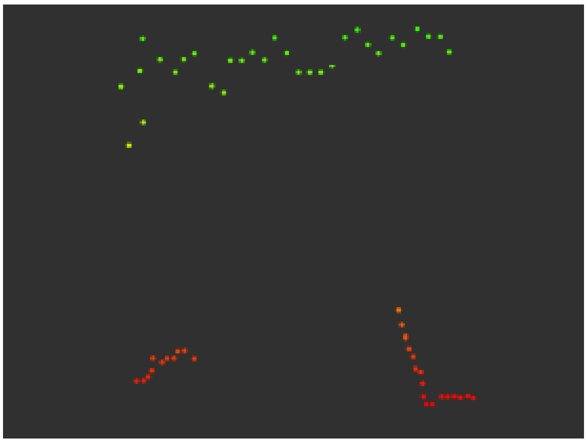


*Figure 8: Aerial view of the resulting nearest point laser scan*

Using a cane does not eliminate the chance of falling, which could occur, for instance, if a person trips when using stairs or if they don't get enough information about where and how a curb is shaped. The requirement to entirely dedicate an arm and hand's function to cane use is another frequent source of annoyance. It can happen when bringing a plate of meals in a cafe or when using a cane and a basket simultaneously in a grocery shop. In densely populated locations with many pedestrians, using a cane presents another problem. Many participants had experienced agitation and reactivity when they unintentionally bumped into or made touch with a sighted person's heel. The frequency of touch with sighted people significantly increases when navigating through dense crowds. Last but not least, many people describe the widespread occurrence of banging their heads on low-hanging objects, such as flowerpots or tree branches.

When asked what kind of technology they prefer there were so many different answers. There were many thoughts about comprehending their environment better. Sometimes, GPS navigation smartphone apps can guide you to a destination, but they cannot convey the existence of nearby details. The existence of vehicles, trees, benches, and other objects piqued people's curiosity. The use of lightweight, compact devices is preferable to those that are large and heavy. Because some people have hearing loss, vibration feedback is especially useful. Others agreed, stating that they would prefer to hear descriptions of the objects' locations and their distances.

*Figure 9: white cane, Xtion Pro Live, EliteBook 850, and Nintendo Wii remotes*

The development of GPS, mobile technology, and other technologies has enhanced blind people's ability to navigate the world. Local navigation, on the other hand, has not made as much progress as technology in terms of object identification, path recognition, and safety. Although there are many electronic mobility aids to buy, the usage of those things by the blind is not very common [26]. The goal is to create an additional obstacle identifying system for the long white cane that overcomes its shortcomings and improves on its advantages. This is achieved by employing sensing techniques and algorithms that support local navigation for the blind using 3D data. The techniques are going to be used in a portable device that helps the blind by giving them more sense about their surroundings. Given the anticipated growth in the number of blind people, the necessity for a device that can assist the blind in efficiently, safely and successfully navigating their environment without falling, tripping, or bumping into barriers is even more critical. According to one organization, population growth and aging might cause the number of blind individuals to quadruple over the following three decades.

# 3. RESEARCH OBJECTIVES

## 3.1 MAIN OBJECTIVE

Real-time object recognition and object access to previously defined classes are the main objective of this component. Through the Smart Navigation system, the user can recognize obstacles, walking surfaces, and potholes using the wearable IoT device and smart cane to provide appropriate information through audio and signal for safe navigation.

## 3.2 SPECIFIC OBJECTIVE

Simply by pointing the camera around themselves, they will be able to identify various items and gauge their size and proximity to them.

1). Capture images using a camera.

a). In proposing system objects detection and classification will be done using MobileNet SSD and deep neural network (DNN) module in OpenCV which detects the object and walking surface and classifies them into different categories.
b). Masked out the background of the obstacle and surface.

2). Following categorization, the object's distance from the camera will be determined.

a). Since numerous objects are recognized at once, the minimum distance from that is taken after calculation. As the user will be hearing about the object and the distance via speech sound, it will cause turmoil for the user.
b). So, the minimum distance object is taken into consideration.

3). We collect various information on them, including their size, name, and proximity to the user.

4). Voice commands are used to convey all messages to the user.

5). The user will receive a warning message if the object is too close to warn them of a potential collision.

# 4. METHODOLOGY

## 4.1 REQUIREMENT GATHERING AND ANALYSIS

- **Conducted a survey**
  After gathering data from internet, a survey was conducted by providing a questionnaire among relatives of visual impaired people to gain an understanding of most important requirements of them to determine the best way to implement the system.

### 4.1.1. Feasibility Study

The feasibility study is useful to determine the proposed project or plan is practical or not

**1. Technical Feasibility**

Project members should have some expert knowledge in web app development and knowledge in software architectures as well as frameworks.

**2. Financial Feasibility**

The proposed sub-component should work perfectly without any errors or failures. The component should be more reliable, high-performance, and less expensive. Limited cost for resources and needs of the component.

**3. Market Feasibility**

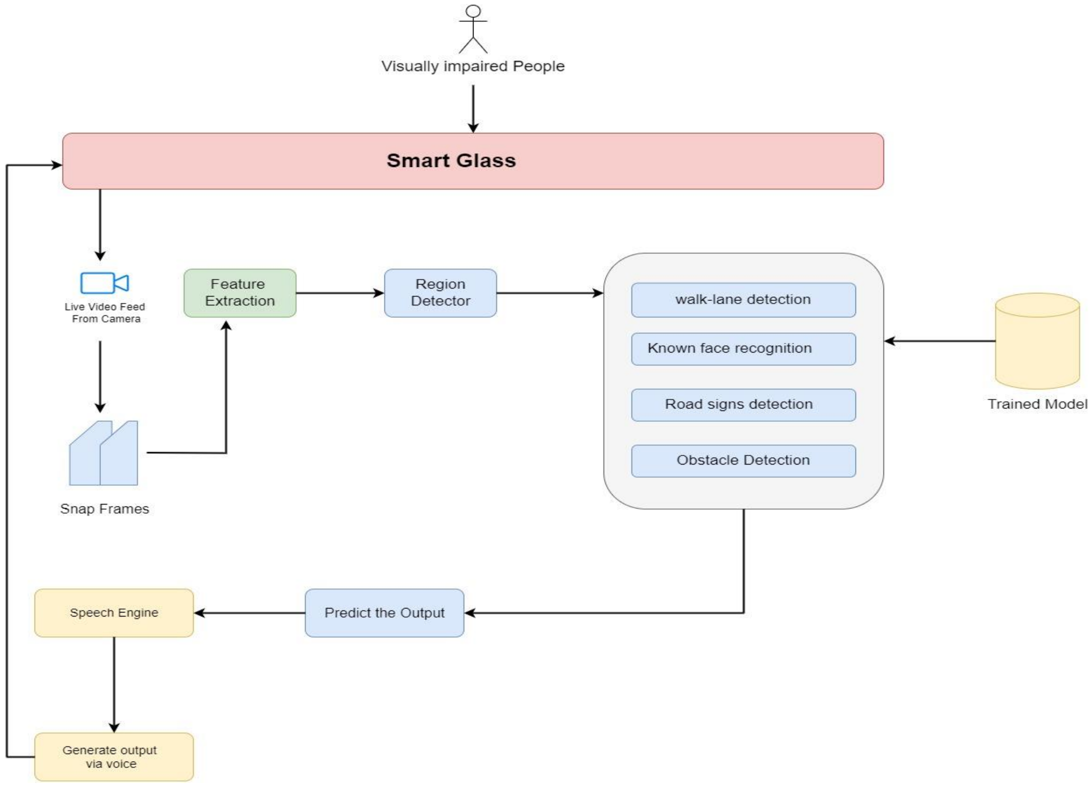
The proposed system should useful product for the industry and provide the best service for the users as well as should have competition with other systems available in the market.

**4. Organizational Feasibility**

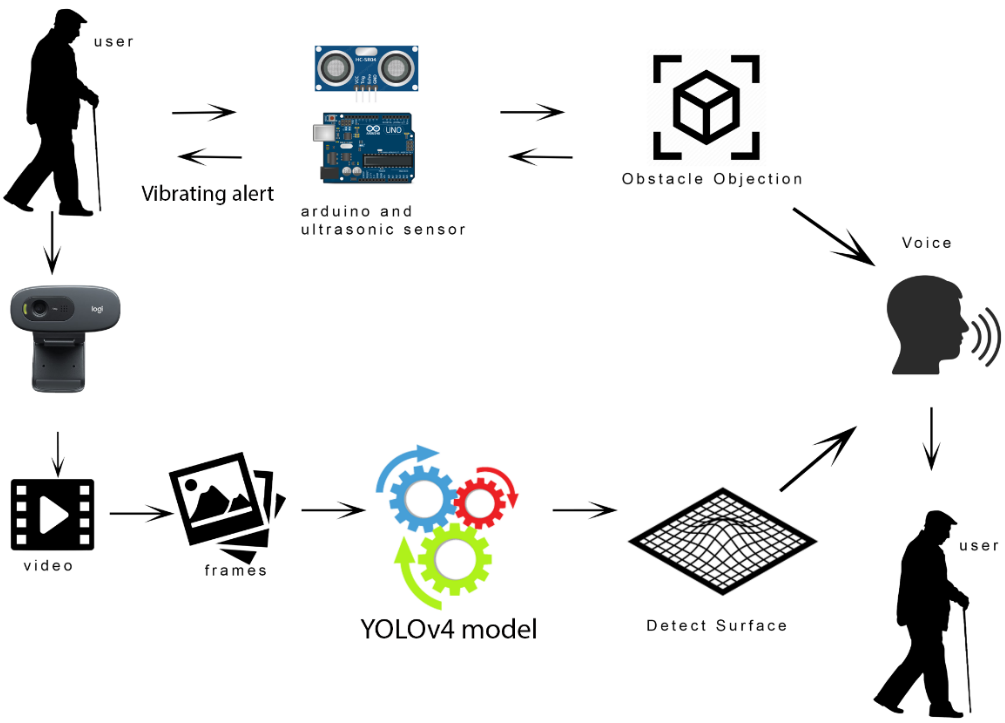
A Member should be responsible for each phase of the software life cycle and especially like requirement analysis phase. The final output of the product should satisfy the identified requirements of the users.

## 4.2. SYSTEM DESIGNS

### 4.2.1. Overall System Diagram




### 4.2.2. Design Diagrams for the component



### IOT technology

IoT is a network of physical objects, or "things," that can connect to and exchange data with other devices and systems through the internet. These "things" can be anything from basic household items to highly developed industrial instruments. Because we can now use embedded devices, low-cost computing, the cloud, big data, analytics, and mobile technologies to connect ordinary objects like vehicles, household appliances, and thermostats to the internet, IoT is crucial today.

### Convolution Neural Network (CNN)

CNN is the best deep learning architecture for image-related problems. It is automatically detecting the crucial characteristics without human oversight. CNN follows convolution and pooling operations and follows parameter sharing [27].

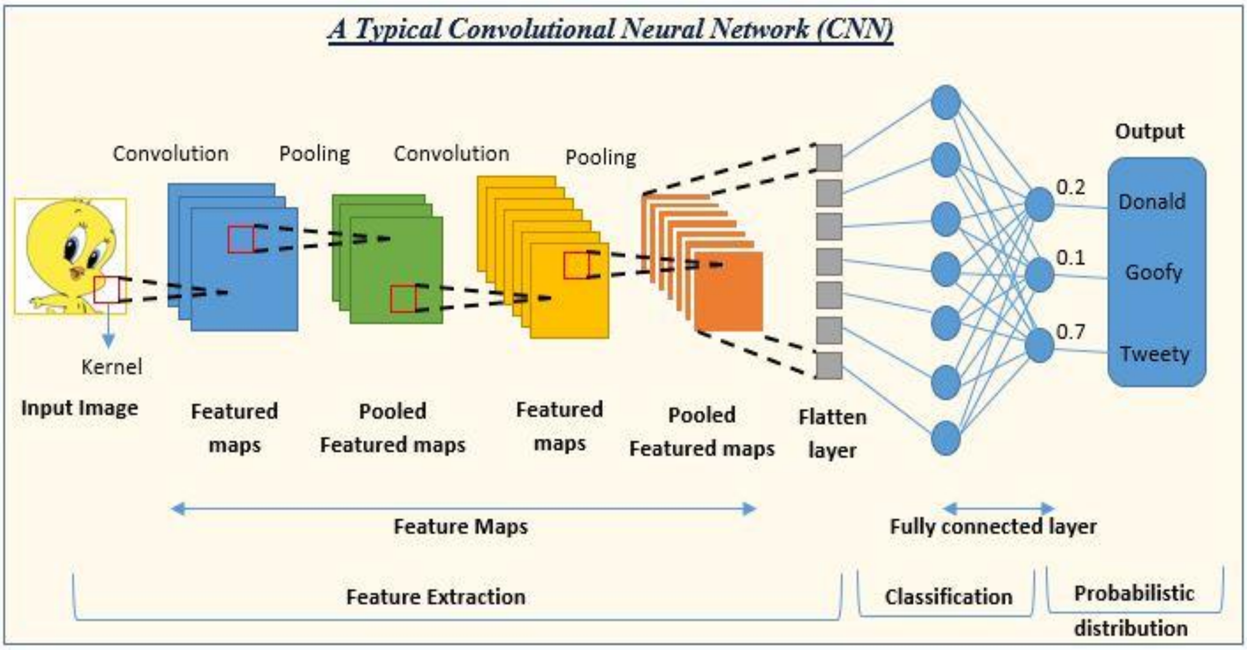



*Figure 10: Common Convolution Natural Network*

Source: https://www.researchgate.net/figure/Convolutional-neural-networkarchitecture_fig7_331733985

As shown in Fig 4.6 Convolutional Neural Network has multiple layers. The convolution layer is the first layer, and it contains filters that are responsible for 25 specific works. Besides, there can be many filters like horizontal, vertical edge filters to determine the edges of the object. When considering the proposed component, when after getting an input image, to get the predictions, which consider as numbers like 0-1 or 0-255 (pixels). Output dimension will change, and input dimensions data also will change, when using convolutions. Thus, better the filter adds a border of zeros and then recalculates the convolution to cover all input values, which process we can call padding. Then convolutional layer, the image of pothole or surface goes through the pooling layer, which has max pooling, min pooling, and average pooling categories. This max-pooling gets a high-intensity value and is the place in the output. In this incident, wherever image can clearly detect and take the information and then put it in the output. Finally, a fully connected layer is responsible to connect all the values of calculations, which the number of layers depends on CNN architectures.

**Machine Learning**

Machine learning has gained popularity recently as a result of the development of computing technology and functions as a subfield of AI and computer science [28]. There is a wealth of data in the world, including audio, visual, and text, and machine learning offers the possibility of extracting meaning from it all. In addition to people, computers, phones, and other technologies also produce data. People examine the data and modify systems to reflect shifting data patterns. However, as the amount of data exceeds our capacity to understand it and manually construct those rules, we will increasingly rely on automated systems that can draw knowledge from the data and incorporate data changes to adapt to a changing environment. Right now, there are numerous immediate uses for machine learning [29].

The primary definition of machine learning is the use of data and providing answers to queries. Making predictions or drawing inferences in order to provide answers is referred to as training and using data is referred to as using data. Training is the process of developing and optimizing a predictive model utilizing data as input. Then, using this predictive model, predictions on never-before-seen data can be made, providing the answers to those queries [29]. The key to this entire process is data, which can be used to refine the model over time as more data is obtained and deploy new prediction models. The secret to machine learning is data, just as the secret to finding that hidden insight in data is machine learning.

**Cloud Platform**

Different types of services are delivered through the Internet. The cloud platform is referring the operating system and hardware of a server. There can be several cloud platforms like public cloud, Private Cloud, Hybrid Cloud which allowed several kinds of benefits. It increases flexibility, improves the scalability and speed, Bandwidth, makes a secure environment[30].

## 4.3. METHODOLOGY

### Object Detection

A computer vision implementation known as object detection enables a system (an algorithm) to make an educated guess as to where specific things are located within a digital scene, such as an image or video. Typically, a bounding box is drawn around the recognized object, making it easier for people to find the object than in raw photos. An object in this context is the image that represents a real-world object (URL). It is a recognizable portion of an image that may be understood as a single entity in image processing[31]. This stands in stark contrast to the common belief that an image or an object can be substituted for one another.

In most cases, an image contains one or more objects, the visibility of which is crucial. An image often contains one or more things, and it's important that they can be seen. Although "detection" may refer to finding a concealed thing, it can also refer to an intelligence's capacity to indicate the existence and identification of an object. It is not necessary to conceal the thing in question.

Many computers vision tasks, including image annotation, activity recognition, face detection, face recognition, and video object co-segmentation, make extensive use of object detection. It is also utilized for object tracking. Some of its many applications include tracking a ball during a football game, the movement of a cricket bat, or a person in a movie.

There are fundamentally two methods for detecting images: machine learning methods and deep learning methods. In more conventional ML-based methods, groupings of pixels that could comprise an object are identified by analyzing various aspects of a picture, such as edges or the color histogram. The location of the object and its label are then predicted using these features as inputs into a regression model. The Scale-invariant feature transform (SIFT), the Histogram of oriented gradients (HOG) features, and the Viola-Jones object detection framework based on Haar features are some examples of machine learning techniques.

In deep learning-based systems, end-to-end, unsupervised object detection is performed using convolutional neural networks (CNNs), which does away with the need for separate feature definition and extraction. Examples of deep learning systems include You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD), Single-Shot Refinement Neural Network for Object Detection (RefineDet), Retina-Net, and Deformable convolutional networks.

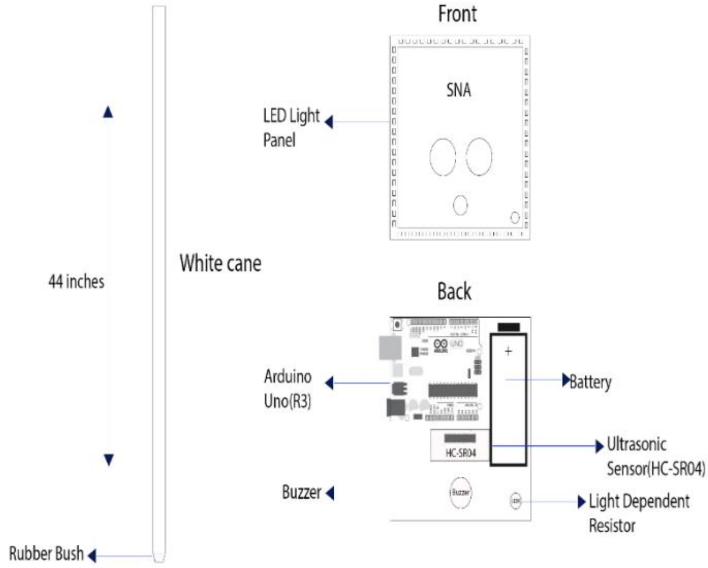



*Figure 11: Blueprint of smart cane plugin*

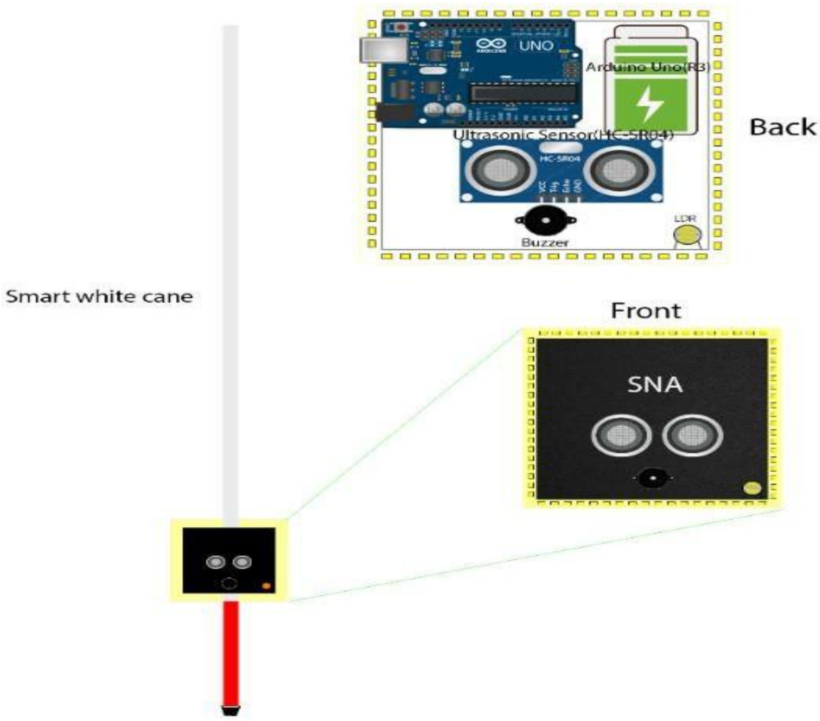



*Figure 12: Prototype diagram for smart cane*

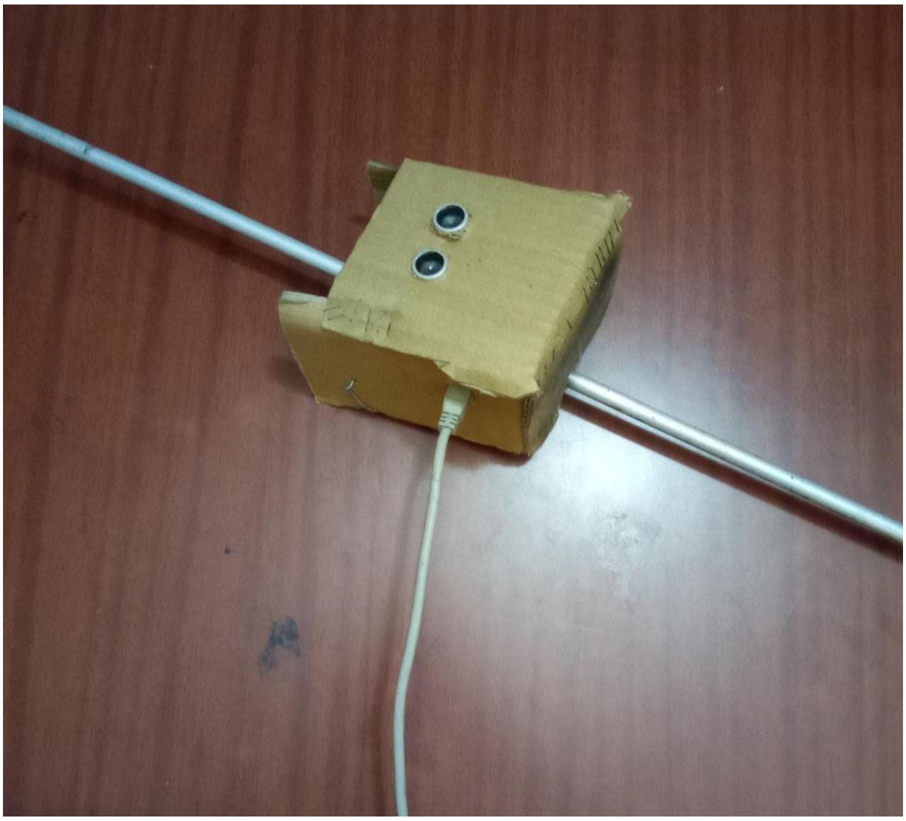


*Figure 13: hard ware*

**Unified Detection Model – YOLO**

Joseph Redmon first introduced the YOLO model in his work "You only glance once, Unified, Real-time object detection." The algorithm's approach uses a single neural network to predict the bounding boxes and class labels for each bounding box directly from the input of a photograph. On speed-optimized versions of the model, it had speeds of up to 45 frames per second and up to 155 frames per person, but it had worse prediction accuracy, primarily due to more localization errors.

To begin, the model divides the input image into a grid of cells. Each cell is then responsible for predicting a bounding box if the center of a bounding box falls within it. Each grid cell has a bounding box that includes an evaluation criterion for quality called a confidence score, the x, y, width, and height. Additionally, a class prediction is based on every cell.

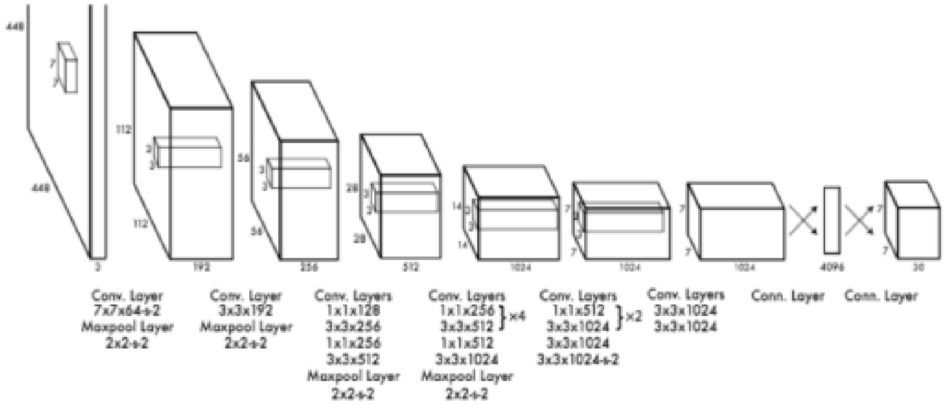


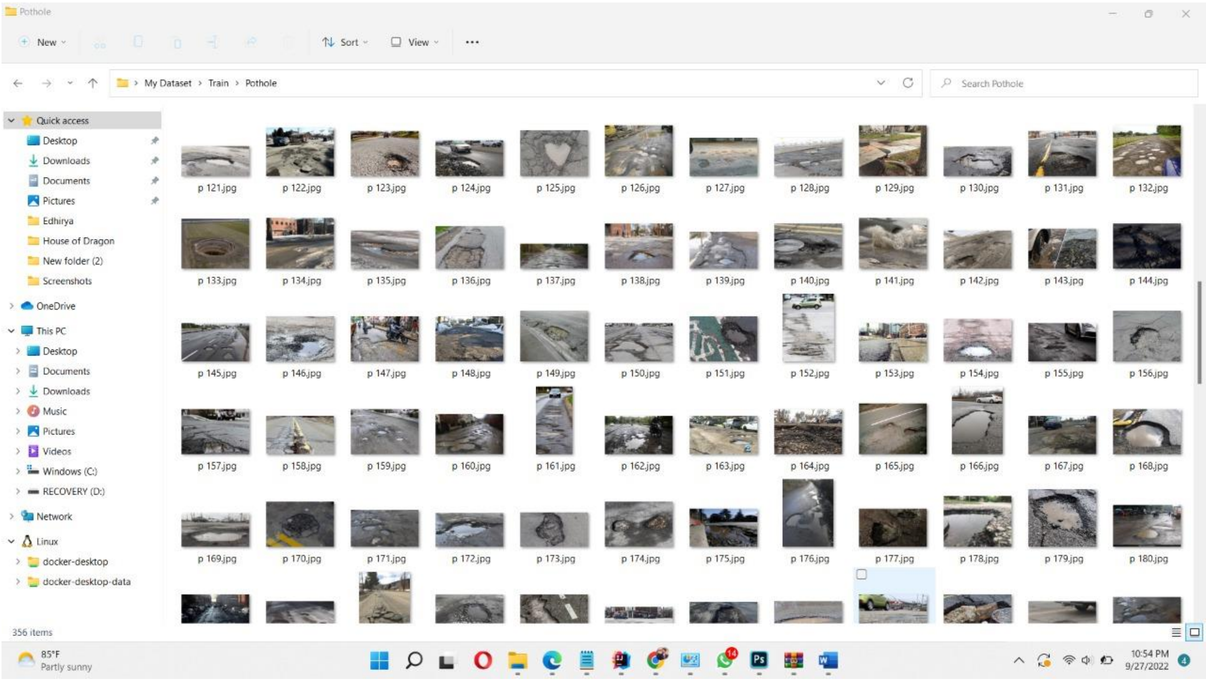

*Figure 14: YOLO Structure*

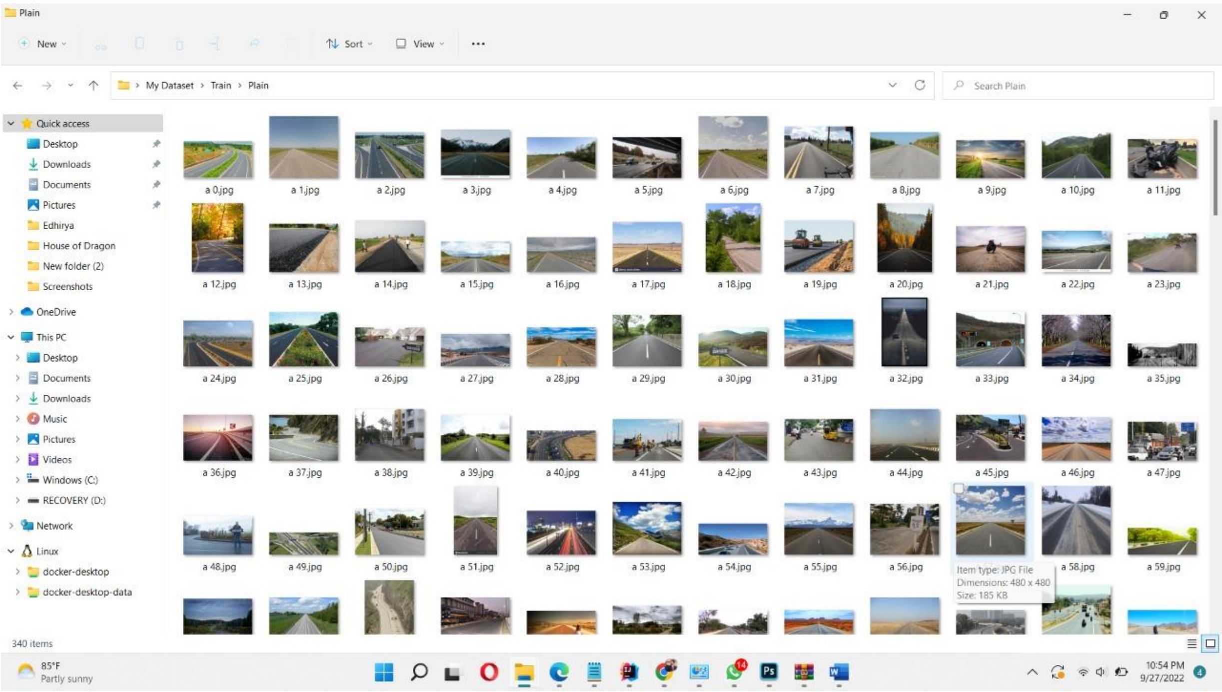
An example will be given to add greater emphasis. A 7 x 7 grid, for instance, might be used to divide a picture into cells that each forecast two bounding boxes, yielding a total of 94 predictions. Combining the bounding boxes with confidences and the map of class probabilities yields the final set of bounding boxes and class labels.

The YOLO was not without flaws; the algorithm had some restrictions on the amount of grids it could operate on, in addition to some other problems that will be discussed later. First, the model employs a 7 7 grid, and because each grid can only identify one object, it caps the number of objects that can be detected at 49. The model also has a problem with near detection, which prevents it from recognizing several objects in a grid cell because each grid can only detect one object at a time. Thirdly, since an object's location could extend outside a grid, there is a chance that the model will mistakenly identify the object more than once. The aforementioned difficulties faced when running YOLO made it quite clear that the system's localization error and other issues needed to be fixed. As a result, YOLOv2 was developed as an enhancement to address the problems and queries raised by its forerunner. As a result, both real and localization problems were heavily rectified in the new 30 version. In their 2016 work "YOLO9000: Better, Faster, Stronger," Joseph Redmon and Ali Farhadi upgraded the model to significantly improve model performance.


*Figure 15: YOLO Network Architecture*

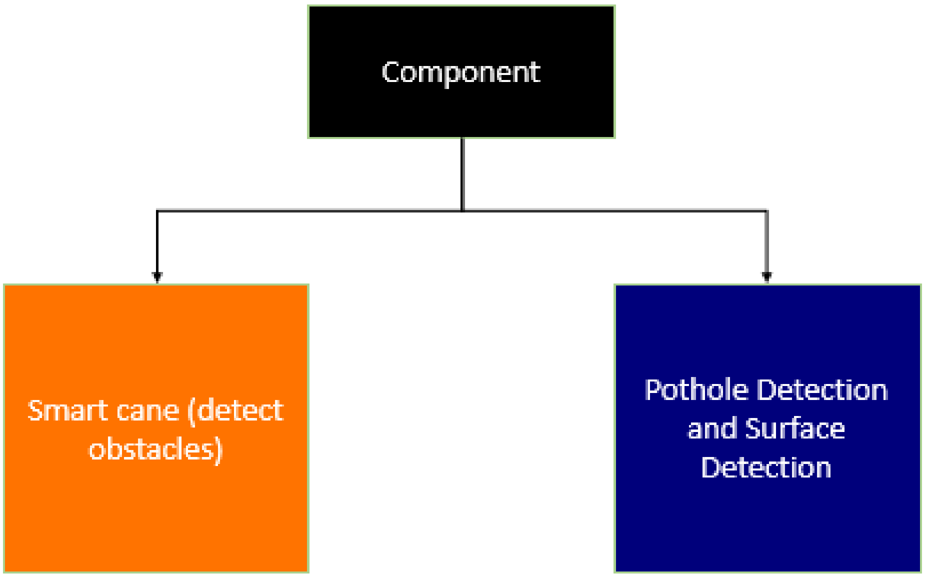
**Datasets**





# 5. IMPLEMENTATION AND TESTING

## 5.1. IMPLEMENTATION




### A. YOLO-based Models

Although YOLOv5 is a single-stage series network with higher accuracy and processing speed, in practice it needs a lot of GPUs for high-volume training. As a result, certain YOLOv5-based network architectures have been developed. The traditional YOLOv3 model is first improved with three different detection scales, an increase in the detection effect on small targets, a modification to the loss function, a logistic regression prediction of the targets' confidence and class, and the use of multi-label classification as opposed to SoftMax classification. The second objective of YOLOv3 small is to increase detection speed. It varies from YOLOv3 in that it fits only two detection layers and minimizes the number of convolutional structures, which dramatically reduces the inference time and decreases accuracy.

SNT is created using the YOLOv5 model, which has a quick response time and effective obstacle detection system. A mobile application is used to detect objects, alerting the user as it does so. For the accurate and quick recognition of objects in urban roadways, the car processing unit recognizes the real-time video of the driver's vision and feeds it to the model. Each frame from the processed input video serves as an input to the object recognition and detection algorithm (YOLOv5). The algorithm divides each frame into three stages: the backbone, neck, and head.

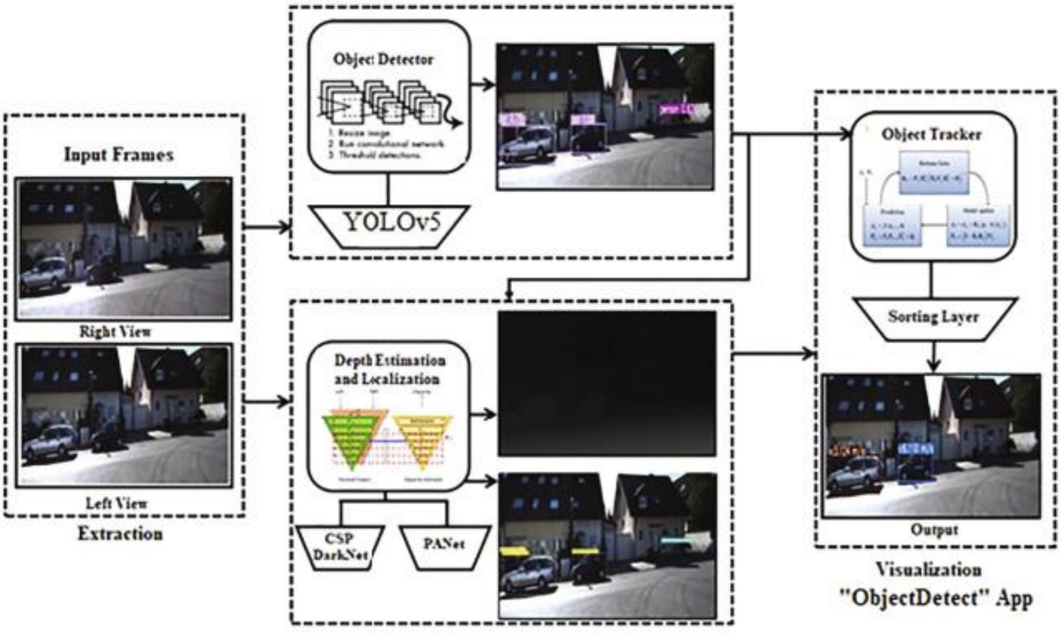



FIGURE 1: Proposed "ObjectDetect" framework for object detection and tracking.

The proposed system is made up of three main elements. This trio of modules is.

(i) Object extraction - backbone and neck
(ii) Object detection and tracking - head
(iii) Object visualization

Object extraction

Using CSPDarknet53 and SPP, PANet path-aggregation, the backbone and neck use images (each of the frames) as input to extract the feature maps. There are 53 convolutional layers in Darknet53. We have 106 layers of architecture for detecting jobs after adding 53 layers to the original 53 levels of architecture.

Step 1: Video input is processed frame by frame in this first step.

Step 2: CSPDarknet53: cross-stage-partial-connections are utilized to reduce redundant gradient data that develops while using traditional DenseNet.

(i) The base layer of CSPDenseNet is separated into portions A and B in this instance

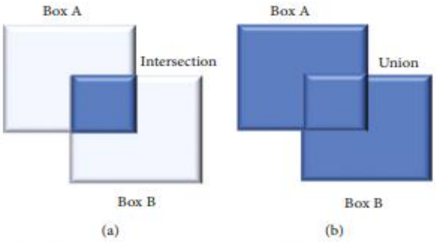



Figure 3: (a) The intersection of bounding boxes. (b) The union of bounding boxes.

(ii) One portion will be treated in the dense block that was created initially; in this case, part B is processed in the dense block.

(iii) The transition stage will be reached without delay in the other section.

As a result, there is no redundant gradient information, and numerous computations are also decreased.

Step 3: The neck is used to add more layers between the backbone and the head. The YOLOv5 method aggregates the data using a modified path aggregation network, together with a modified spatial attention module and a modified SPP. The accuracy of the detector is improved by using concatenated path aggregation networks with spatial pyramid pooling (SPP) extra modules.

Object Detection

Each frame that has been processed in the neck and backbone is then transported to the head using the YOLOv5 algorithm, which employs the following methods:

Step 1. - Grids are originally used to partition the input frame into residual blocks. Object presence detection falls under the purview of each grid cell.

Step 2 of the YOLO method involves projecting bounding boxes and confidence scores around each object that is present in that particular grid. The bounding box center (x, y), width (bw), height (bh), and confidence score are the attributes that make up each bounding box (c). The confidence score reflects the algorithm's level of assurance and precision for a specific object within that bounding box. In addition to these characteristics, YOLO employs a single bounding box regression to forecast the likelihood that an object will appear in the bounding box. On a camera, the YOLOv5 algorithm is being used in real time. By identifying the classes to which the items belong and the confidence scores that indicate how certain it is that the objects are present, the program identified objects in the frames.

Step 3: Find the intersection over union (IoU) between the predicted box and the ground truth. If there is no item in a grid cell, the confidence score is zero; otherwise, it must equal the intersection over union (IoU) value. Step 3: Find the intersection over union (IoU) between the predicted box and the ground truth. If there is no item in a grid cell, the confidence score is zero; otherwise, it must equal the intersection over union (IoU) value. Since the user has personally set the ground truth boxes in this case, a higher IoU translates into a higher confidence score, which raises the algorithm's prediction accuracy. Based on the likelihood that the box contains items, those without objects are filtered out. Nonmax suppression techniques remove undesirable bounding boxes, leaving just the box with the highest confidence or probability score.

$$\text{IoU} = \frac{\text{Area of } (\text{BoxA} \cap \text{BoxB})}{\text{Area of } (\text{BoxA } U \text{ BoxB})}.$$

The overlap between two proposals is determined using the aforementioned Equation IoU calculation. Nonmax suppression is used to identify the correct bounding box from among those that have been predicted by algorithms based on confidence scores. Algorithm 1 below illustrates this.

The method finds the object and class probabilities with confidence ratings in Step 4 of the detection process.

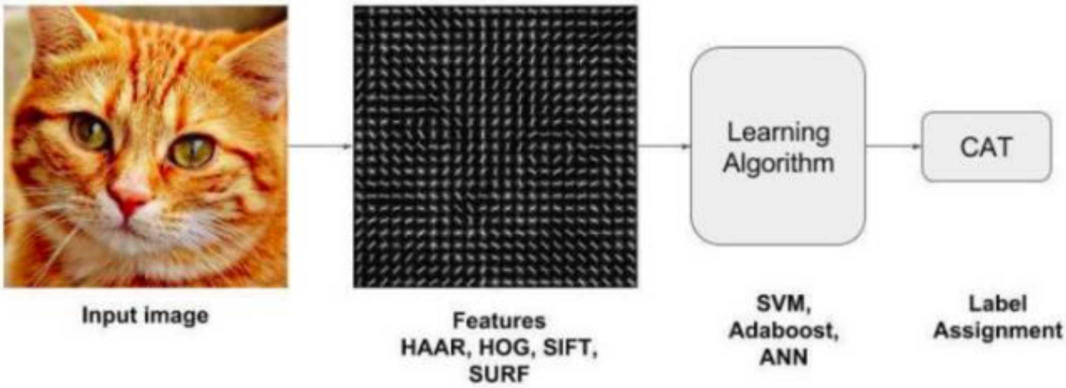
**Visualization**

Finding and locating items in pictures or videos is what object detection includes. The process of object detection encompasses a variety of techniques, including, but not limited to, object annotation, image preprocessing, bounding box localization, and image classifications. Object detection is widely used in industrial processes, personal gadgets, and public services. Bounding box detection on your phone is one of the key use cases that you encounter on a daily basis.

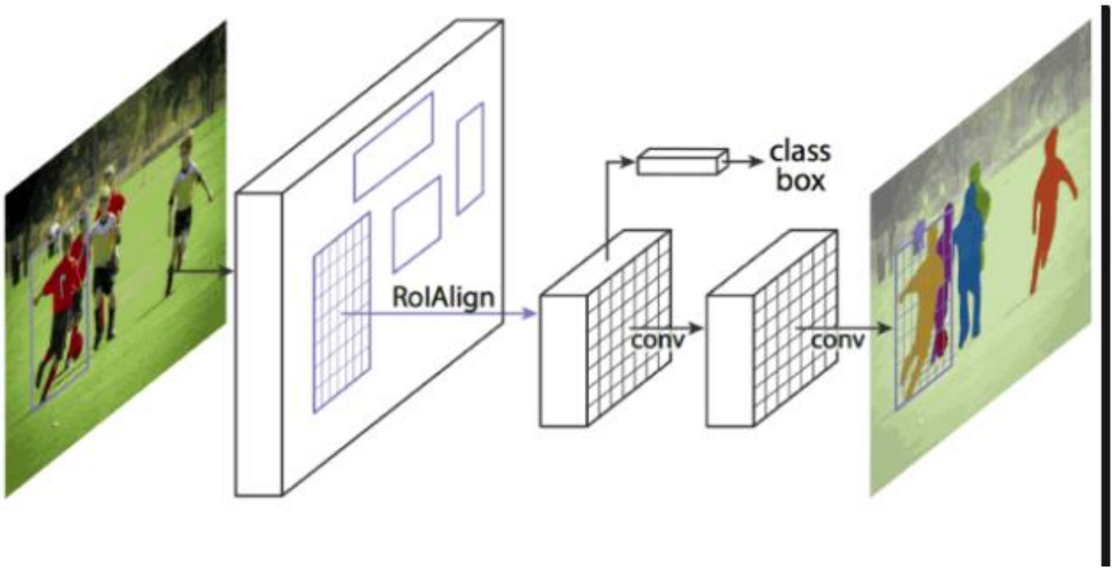
From such straightforward use cases as face detection, object detection techniques can be applied to real-world situations with significant effects, such as preventing traffic accidents, finding flaws in industrial assembly lines, detecting for military purposes, etc.

Detecting objects is not a brand-new phenomenon. It has been around since the dawn of computer vision and has advanced significantly from the previous methods. In the early stages of object detection, features were manually extracted before items were defected using classifiers.

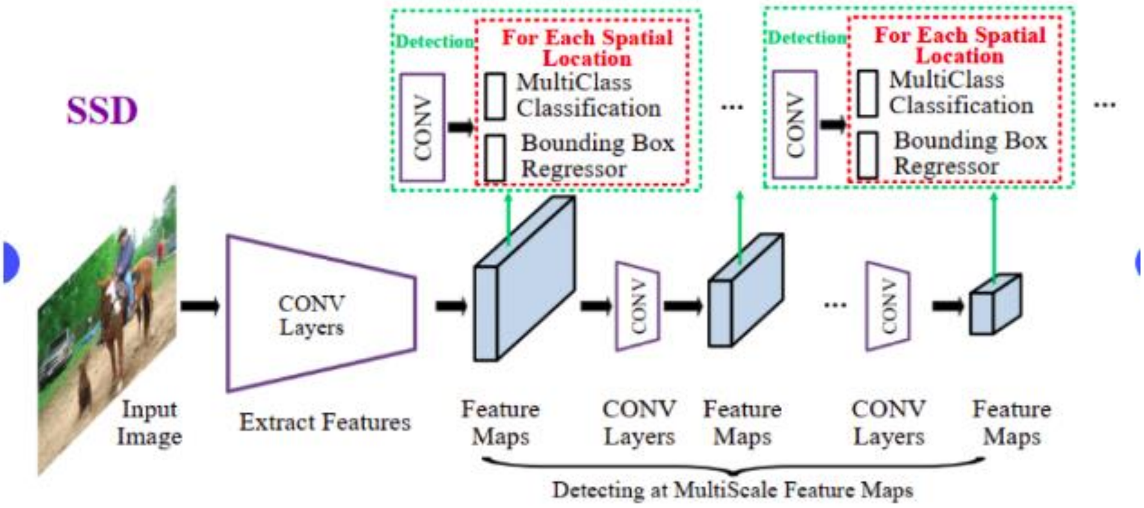
The HOG (histogram of oriented gradient), Haar, and SIFT approaches were some of the first methods for feature extraction ( scale-invariant feature transform). Following the extraction of features using these algorithms, the classification of the images was carried out using SVM (Support vector machine) or other classification algorithms like Random forest or Adaboost.



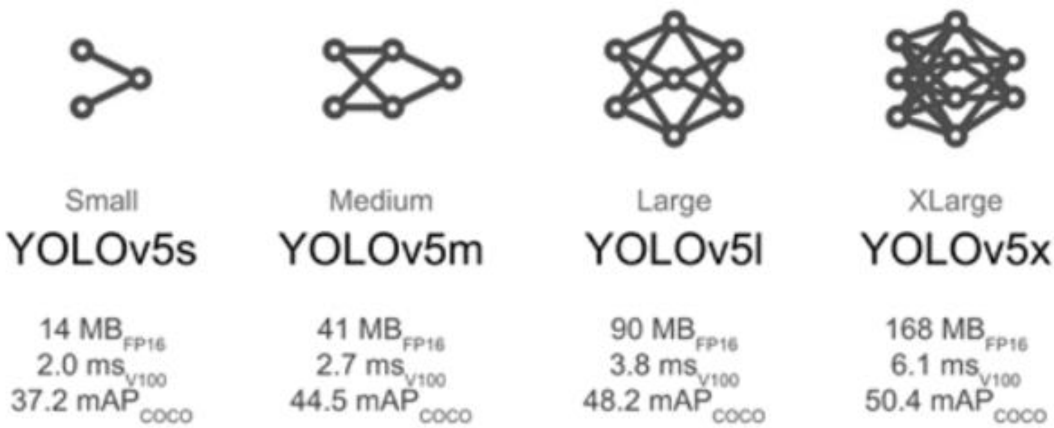

Traditional machine learning methods were not scalable for complicated use cases because they relied on the extraction and categorization of low-level feature information. But as deep learning became more widespread, a number of algorithms became more refined and scalable to a variety of application scenarios. R-CNN (region-based convolutional neural networks), SSD (single shot multiBox detection), and YOLO are a few of the well-known ones ( you only look once). Compared to the conventional procedures, these techniques offered substantially higher accuracy. These days, numerous frameworks, including Tensorflow and Pytorch, offer capabilities for bespoke object detection. For jobs involving object identification, some of the well-known frameworks, like transformers, are also commonly employed. The Vision Transformer is one such object detector used in picture classification.



class
box
RoIAlign
conv
conv

SSD
Detection
For Each Spatial Location
MultiClass Classification
Bounding Box Regressor
CONV
CONV Layers
Input Image
Extract Features
Feature Maps
CONV Layers
Feature Maps
CONV Layers
Feature Maps
Detecting at MultiScale Feature Maps

Small
YOLOv5s
14 MB FP16
2.0 ms V100
37.2 mAP COCO
Medium
YOLOv5m
41 MB FP16
2.7 ms V100
44.5 mAP COCO
Large
YOLOv5l
90 MB FP16
3.8 ms V100
48.2 mAP COCO
XLarge
YOLOv5x
168 MB FP16
6.1 ms V100
50.4 mAP COCO

We will examine the development of many strategies in this post and gain a conceptual understanding of them. This article will establish the groundwork for the object detection program that we will gradually develop throughout this series. We will gain practical expertise with each of these strategies throughout this series before tying them all together in the pothole detection application, where we will utilize the trained model to identify potholes in videos.

**Template matching**

The template matching method for object detection in an image might be described as naive. This method involves dragging a template of the target object across the image and taking a picture of how well it matches the input image. The projected location of the object is the area where the correlation is greatest.


The pothole template is slid across the image, as seen in the figure above. The best matching site is found using the correlation coefficient between the template and image's pixel intensities. Implementing this is simple using frameworks like OpenCV. Template matching is a straightforward method that has a number of drawbacks. The issue with using different scales for the template and the image is one that frequently arises. The detection of objects will frequently be incorrect if the template and image scales mismatch.

The visual variation of the object in the template and image is another issue. Object detection on the image degrades greatly if the visual effects of the objects differ from those of the template.

One of the earliest approaches used for object detection was template matching, however it is no longer used by any of the modern object detectors.

**Image pyramid and sliding window methods for Object detection**

Assume that the window used to detect potholes has a set size and that a pothole is only identified if it properly fits inside the box. We might not be able to identify every pothole that might be present in an image with such a fixed-sized window. While the larger potholes at the near end of the image are not recognized since the box is smaller than the pothole, it is clear that the fixed-sized window was able to identify one of the potholes further down the road because it fit well inside the window size.

Let's gradually shrink the image size while maintaining the box's size as a solution to this problem. As we move from layer 1 to layer 7 in the image below, this can be seen. Since our detection window doesn't change, we are able to detect potholes of all sizes as the image size of the item we wish to detect decreases.

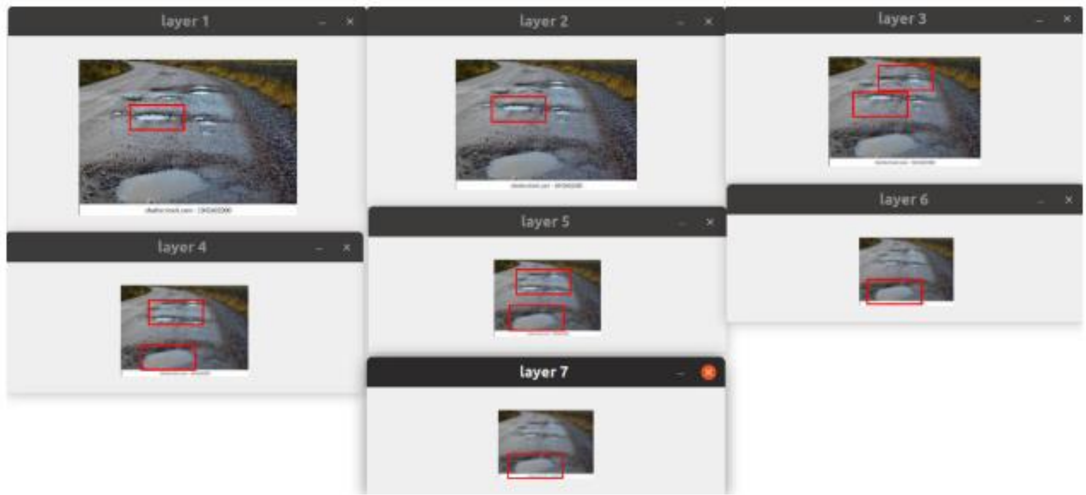



*Figure 16: Object detection with fixed size window*

The fundamental method in picture pyramids is the practice of gradually scaling an image to find things. The term "picture pyramids" refers to the fact that scaled images can fit inside a pyramid when stacked vertically, as seen in the figure below.

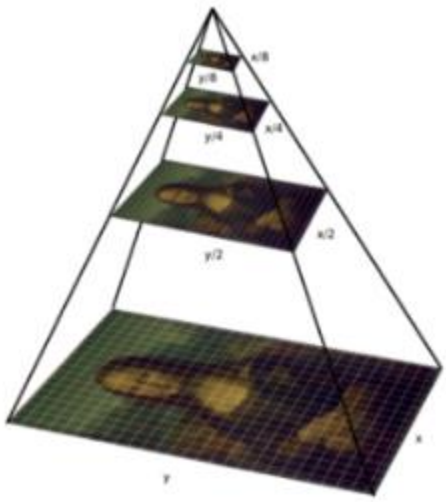



*Figure 17: Image Pyramids*

The image pyramid can be used in many different ways. Laplacian and Gaussian pyramids are two of the more notable ones.

Pyramidal images by themselves cannot identify items. This technique must be used in conjunction with a technique known as sliding windows that enables item detection in a picture at different scales and locations. As the name implies, this technique involves extracting information from an image by drawing a window of standardized length and width over it. The object of interest will then be identified using a classifier that incorporates these qualities.

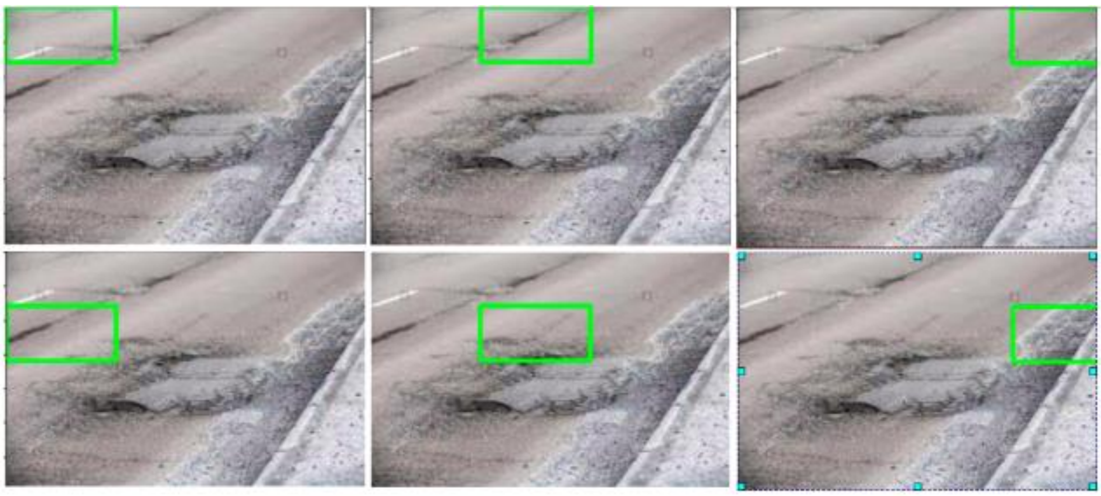
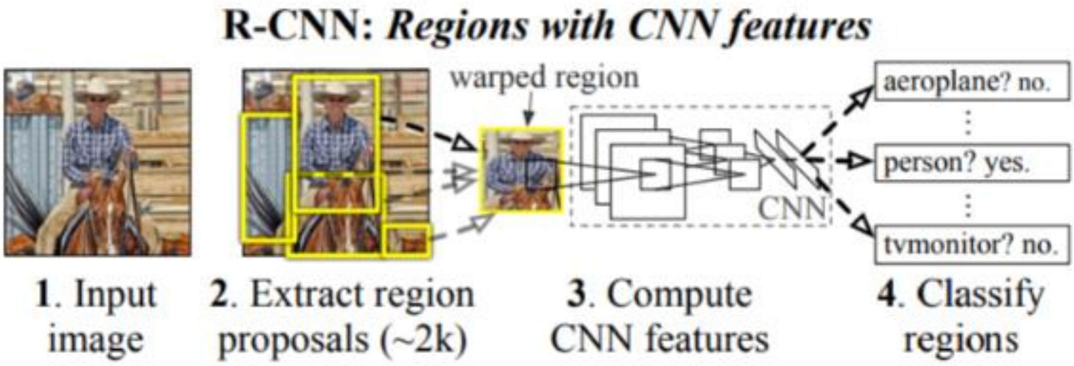


Figure 18: Sliding window accross the image to detect objects

**RCNN Framework**

The original architecture has undergone numerous revisions, which have improved performance over time. The RCNN framework was the preferred model for object detection tasks for a while.




The following crucial steps make up the original RCNN algorithm.

- Extract from the input image all areas that might contain an object. Regions proposed extractions are what these extractions are known as. A method of extraction known as selective search was employed.
- To extract features from the proposal regions, use a pretrained CNN.
- Sort out each extracted region using a classifier such as Support Vector Machines ( SVM).

Compared to conventional techniques like the sliding window and pyramid-based algorithms, the original RCNN algorithm produced far better results. But this system moved slowly. Additionally, deep learning was not employed to localize the items in the image; instead, algorithms like selective search were mostly utilized.

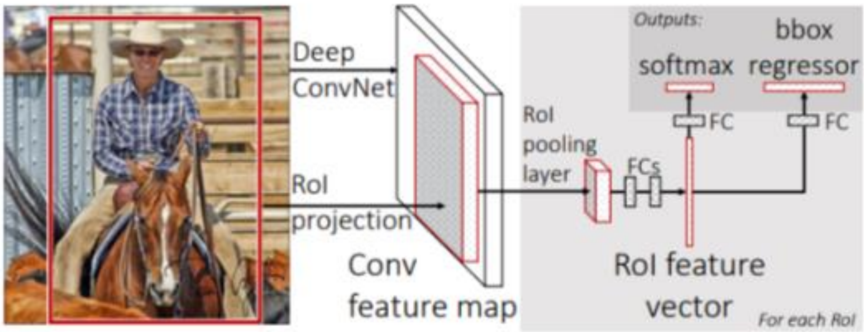



Figure 19: Fast-RCNN Architecture

Within a year after the initial paper's publication, the same author significantly improved the original RCNN algorithm. The Fast-RCNN algorithm was used. This method contained some original concepts, such as the Region of Interest Pooling layer. In order to extract a feature map from the complete image, the Fast-RCNN method uses a CNN. To obtain the output label for the proposal regions, a fixed size window from the feature map was taken and then sent to a fully connected layer. The Region of Interest Pooling was the name given to this procedure. The class labels of the regions and the locations of their bounding boxes were obtained using two sets of fully connected layers.

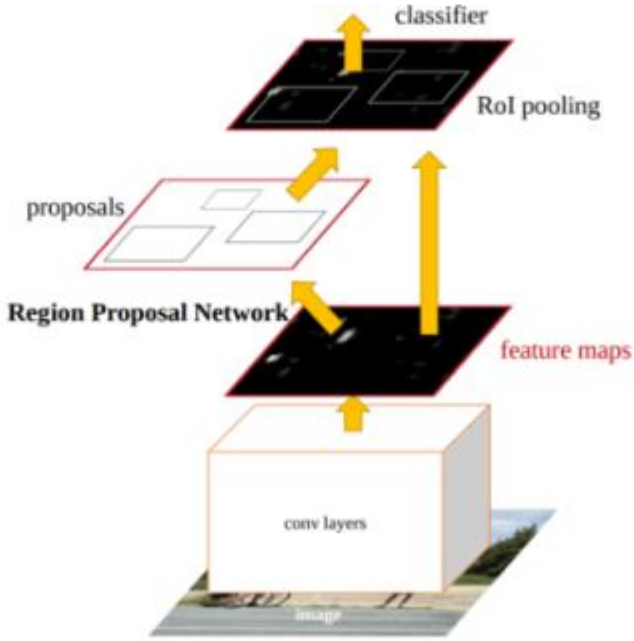



*Figure 20: Faster RCNN*

The Faster-RCNN algorithm, which was announced a few months after the Fast-RCNN algorithm, was an improvement over the earlier method. The Region Proposal Network (RPN), a key component of the new method, was designed to do away with the necessity for selective search algorithms and incorporate region proposal functionality directly into the R-CNN architecture. The anchors used in this technique were distributed unevenly over the entire image at various scales and aspect ratios. The RPN would assess these anchors and output a suggestion as to where an object is likely to exist.

An image's possible bounding box regions are generated by the R-CNN architecture. After that, a classifier is used to classify these prospective locations. Following classification, these pre-processed regions are used to improve bounding boxes, get rid of redundant detections, and rescore boxes on other items in the image. These pre-processed regions are then utilized to enhance bounding boxes, eliminate redundant detections, and rescore boxes on additional objects in the image after categorization.

**YOLO Algorithm**

A straightforward approach known as YOLO—an acronym for "You Only Look Once"—treats the issue of object detection as a single regression problem, processing an image straight through from pixels to bounding coordinates and class probabilities. A single neural network at the heart of this approach predicts several bounding boxes and class probabilities concurrently.

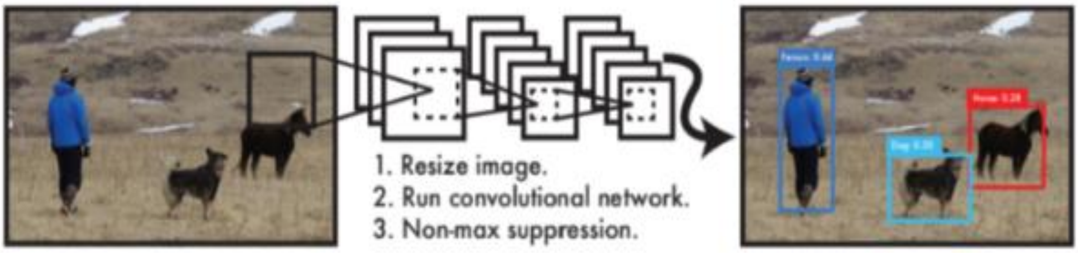
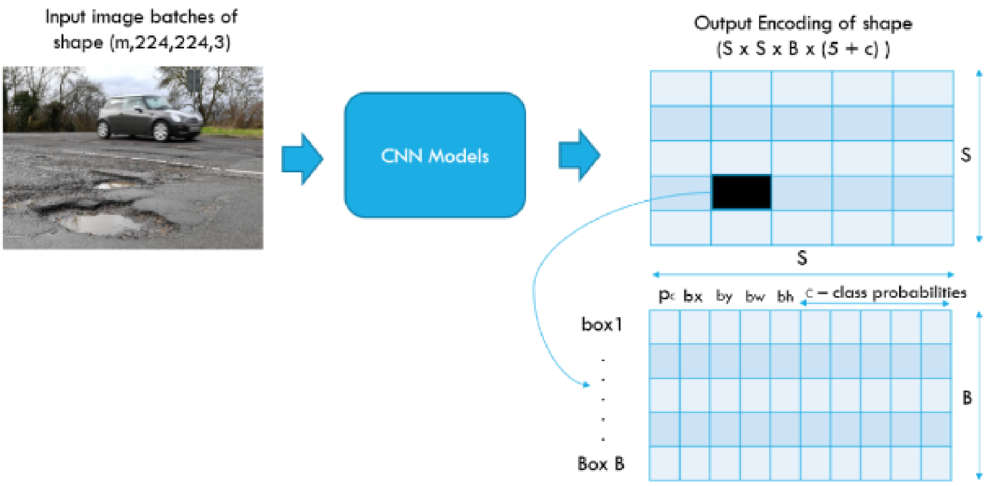



*Figure 21: Yolo Algorithm*

An input image is separated into grids of identically sized squares. Multiple bounding boxes are predicted by each grid, together with a confidence rating for each box. The confidence scores show how certain the model is that an object is in the box. By combining all of the object detection components into a single neural network, YOLO can anticipate each bounding box by using characteristics from the entire image. All classes' bounding boxes are predicted concurrently. YOLO comes in a variety of forms, from YOLOv1 to YOLOv5, to the PP-YOLOv2 that was released in April 2021. The models' accuracy is comparable to but typically not superior than R-CNNs; however, where they differ is in their speed of detection, which makes it an excellent choice for real-time video or with camera feed. In the fifth article of this series, we will put a YOLO object detector into practice.

One of the most common item detectors in use goes by the acronym YOLO, which stands for "You only look once." The technique is set up so that the network may produce predictions after just one run of forward propagation. YOLO has excellent real-time detection performance and very high accuracy. YOLO provides a list of bounding boxes together with confidence scores and class labels after taking a batch of photographs of the form (m, 224,224,3). (pc,bx,by,bw,bh,c).



The result will be a collection of S x S grids, each with a set of B anchor boxes (for example, 19 × 19). Each box will have 5 fundamental dimensions, including 4 bounding box details and a confidence score. Each box will also contain the probability of the classes in addition to these five essential pieces of information. Therefore, each box will contain a total of 15 cells (5 + 10) if there are 10 classes. YOLO begins the process by dividing the image into S x S grids. In this case, S may be any integer value. Let's assume s is 4 in our case.

A confidence score would be used to predict B boxes for each cell. Once more, B can be determined based on the maximum number of items that can fit inside a cell. The box's center needs to be inside the cell, which is a crucial requirement. The anchor boxes are known as these B boxes.

In our situation, let's assume that B equals 2. Therefore, each cell will anticipate two boxes where there is a chance that an object may appear. Let's look at the grid in the preceding image where two boxes are foreseen. We can see that the pothole and the car were both picked up by that cell, and we can also see that the center of the boxes is also in that cell. Each cell in the image is subject to the box prediction procedure. This stage involves predicting several overlapping boxes across all of the image's grids.

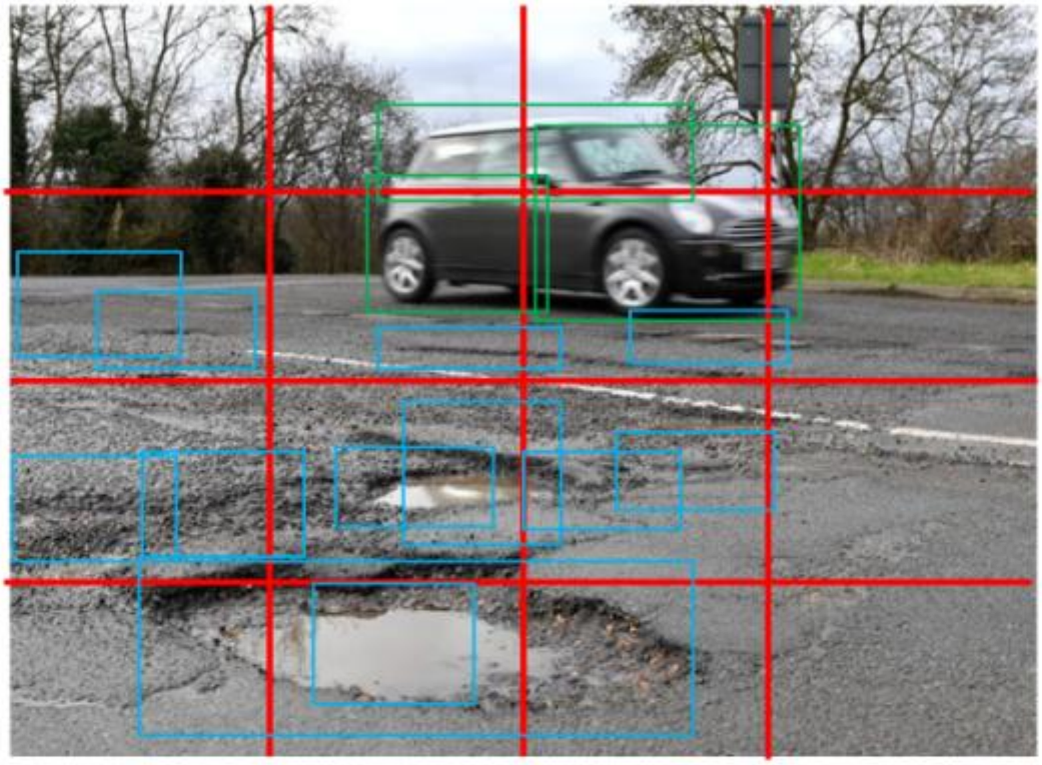


*Figure 22: Generate B boundibg boxes and confidence scores.*

A class probability map is additionally forecasted in addition to the boxes and confidence ratings. A class probability map indicates the possibility of each class being present in each cell. Vehicle, for instance, in cells 2, 3, and 4. and a crack in cells 9, 10, 11, etc.

*Figure 23: Get class probability maps*

The network can assign a class map to each of the enclosing boxes thanks to the class probability maps. To limit the number of overlapping boxes and obtain the bounding boxes for only the items we wish to classify, non maxima suppression is then employed.

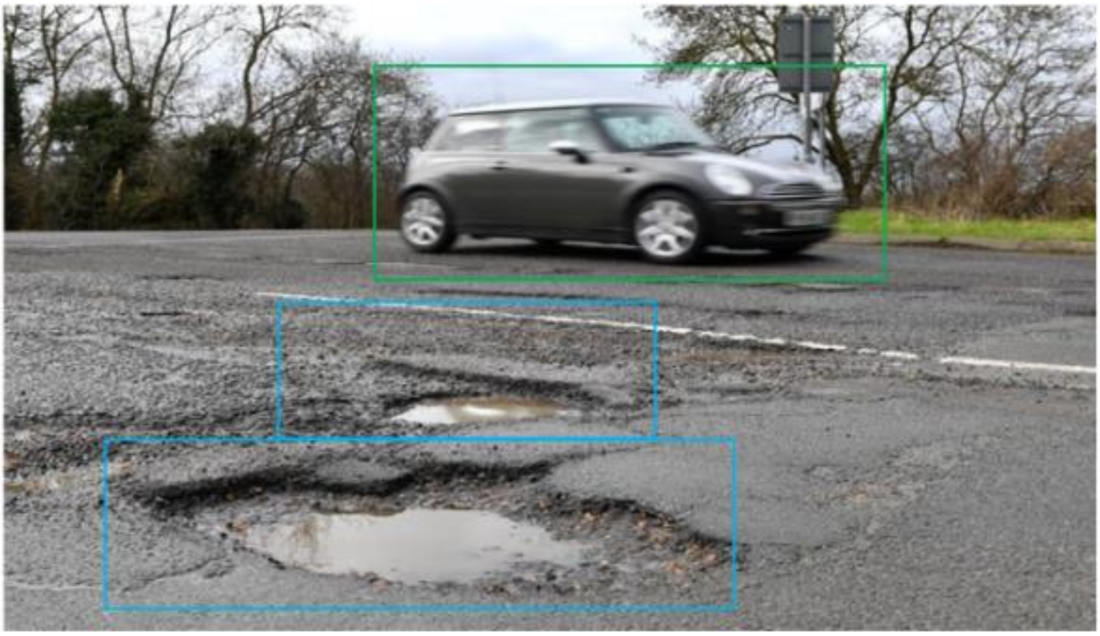
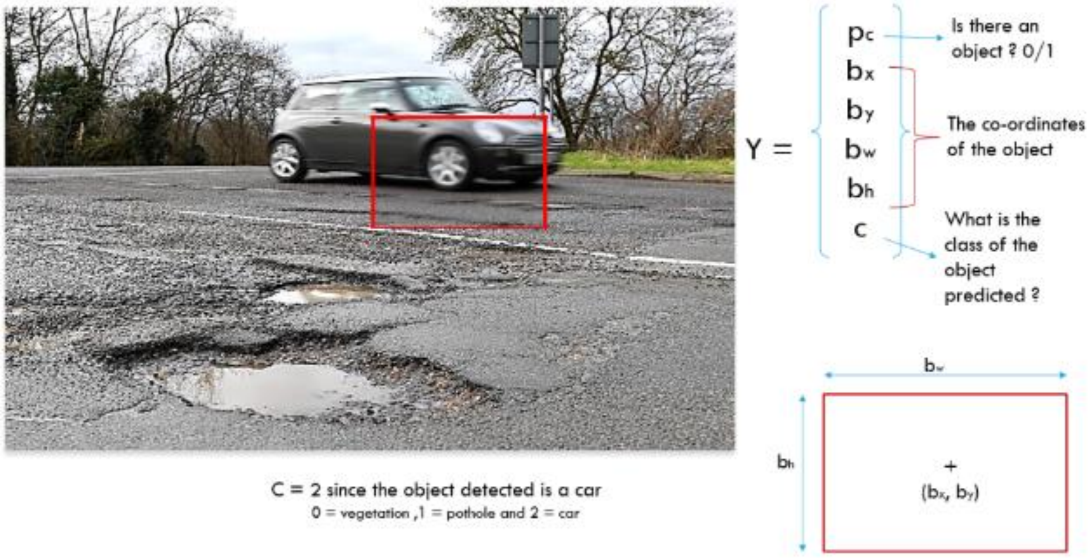


*Figure 24:Final detection after NMS*

After viewing a summary of the entire procedure, let's examine the results or projections from each cell. Let's focus on the particular cell in the image below.




A confidence score is predicted for each cell, indicating whether or not the cell contains an object. The object's bounding boxes and class are also predicted in addition to the confidence score. The class label could be a one-hot encoding representation of the projected class or an integer like 2 or 1. ( eg. [0,0,1] ). Now that you know how YOLO works, let's talk about how it's implemented.

**Implementation of YOLO-V5**

The procedure will be controlled by a Jupyter notebook. Since this model has already been trained, there won't be many processes we need to regulate. The following steps make up the entire implementation procedure.

1. Downloading the model files for the YOLO V5.
2. Preparing the files with annotations.
3. The train, validation, and test sets are being prepared.
4. Putting the training process into action.
5. Employing the trained model to carry out the inference process.

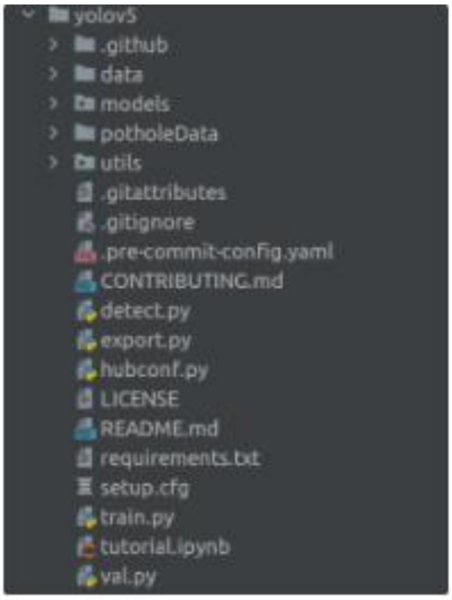
Pytorch was used to train the unique Yolo model. Importing all the necessary packages was the first step.

```
import pandas as pd
import os
import glob
from PIL import Image, ImageDraw
import numpy as np
import matplotlib.pyplot as plt
import random
from sklearn.model_selection import train_test_split
import shutil
import torch
from IPython.display import Image  # for displaying images
import os
import random
import shutil
import PIL
```

The first step is cloning the YOLOV5 official repository. Then clone the repository from Jupyter notebook.

```
! git clone https://github.com/ultralytics/yolov5
```

There are many more default folders under the Yolov5 folder. The folder hierarchy was like the one shown below.




Prepare annotation file

Annotation csv file was used to prepare the annotated file.

```
# Reading the csv file
pothole_df = pd.read_csv('BayesianQuest/Pothole/pothole_df.csv')
pothole_df.head()
```

| | filename | width | height | class | xmin | ymin | xmax | ymax |
|---|---|---|---|---|---|---|---|---|
| 0 | pothole1.jpeg | 275 | 183 | pothole | 64 | 78 | 130 | 107 |
| 1 | pothole1.jpeg | 275 | 183 | pothole | 44 | 105 | 131 | 154 |
| 2 | pothole1.jpeg | 275 | 183 | pothole | 12 | 151 | 59 | 177 |
| 3 | pothole1.jpeg | 275 | 183 | vegetation | 163 | 33 | 254 | 58 |
| 4 | pothole1.jpeg | 275 | 183 | pothole | 115 | 54 | 142 | 74 |

We will now establish a class map, which is a dictionary that associates an integer value with each of our classes.

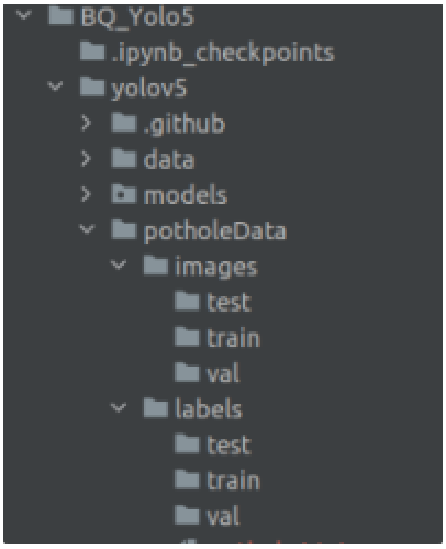
```
# First get the list of all classes
classes = pothole_df['class'].unique().tolist()
# Create a dictionary for storing class to ID mapping
classMap = {}

for i,cls in enumerate(classes):
    # Map a class name to an integet ID
    classMap[cls] = i

classMap
```

```
pothole': 0, 'vegetation': 1, 'sign': 2, 'vehicle': 3}
```

The bounding box information for the photos will then be extracted from the excel sheet in a specific format that YoloV5 requires. Additionally, we must keep the labels and annotation files (pictures) in their respective directories. Before we extract the bounding box data, let's build the folders.

```
# Create the main data folder
!mkdir potholeData
# Create images and labels data folders
!mkdir potholeData/images
!mkdir potholeData/labels
# Create train,val and test data folders for both images and label:
!mkdir potholeData/images/train potholeData/images/val potholeData,
```

The folder structure appears as follows after the establishment of these folders.

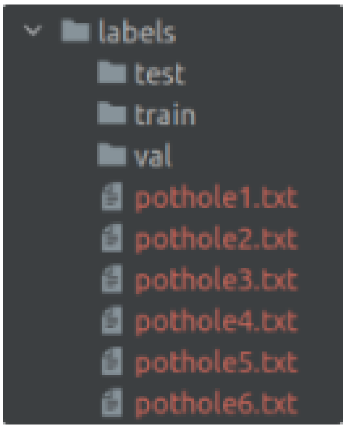



Extracting the bounding box information.

To accomplish that, it must repeatedly run over all of the photos it contains before obtaining the boundary data in the.txt format required by YoloV5.

```
# Creating the list of images from the excel sheet
imgs = pothole_df['filename'].unique().tolist()
# Loop through each of the image
for img in imgs:
    boundingDetails = []
    # First get the bounding box information for a particular imag
    boundingInfo = pothole_df.loc[pothole_df.filename == img,:]
    # Loop through each row of the details
    for idx, row in boundingInfo.iterrows():
        # Get the class Id for the row
        class_id = classMap[row["class"]]
        # Convert the bounding box info into the format for YOLOV5
        # Get the width
        bb_width = row['xmax'] - row['xmin']
        # Get the height
        bb_height = row['ymax'] - row['ymin']
        # Get the centre coordinates
        bb_xcentre = (row['xmin'] + row['xmax'])/2
        bb_ycentre = (row['ymin'] + row['ymax'])/2
        # Normalise the coordinates by diving by width and height
        bb_xcentre /= row['width']
        bb_ycentre /= row['height']
        bb_width    /= row['width']
        bb_height   /= row['height']
        # Append details in the list
        boundingDetails.append("{} {:.3f} {:.3f} {:.3f} {:.3f}".fo
    # Create the file name to save this info
    file_name = os.path.join("potholeData/labels", img.split(".")[
    # Save the annotation to disk
    print("\n".join(boundingDetails), file= open(file_name, "w"))
```

After completing this step, the annotations are visible as text files in the labels folder.




**Preparing the train, test and validation sets.**

All of the train, test, and validation photos, as well as the annotation text files, were required in the corresponding folders that were created in order to train the Yolo model. This stage listed the locations of the images and the texts that were annotated, divided the locations into training, testing, and validation sets, and then copied the image and annotation files to the appropriate folders.

```
# Get the list of all annotations
annotations = glob.glob('potholeData/labels' + '/*.txt')
annotations
```

```
'potholeData/labels/pothole1.txt',
'potholeData/labels/pothole10.txt',
'potholeData/labels/pothole11.txt',
'potholeData/labels/pothole12.txt',
'potholeData/labels/pothole13.txt',
'potholeData/labels/pothole14.txt',
'potholeData/labels/pothole15.txt',
'potholeData/labels/pothole16.txt',
'potholeData/labels/pothole17.txt',
'potholeData/labels/pothole18.txt',
'potholeData/labels/pothole2.txt',
'potholeData/labels/pothole3.txt',
'potholeData/labels/pothole4.txt',
'potholeData/labels/pothole5.txt',
'potholeData/labels/pothole6.txt',
'potholeData/labels/pothole7.txt',
'potholeData/labels/pothole8.txt',
'potholeData/labels/pothole9.txt']
```

```
# Get the list of images from its folder
imagePath = '/media/acer/7DC832E057A5BDB1/JMJTL/Tomslabs/Baye
images = glob.glob(imagePath + '/*.jpeg')
images
```

The photos and annotation files were then sorted, and the data was divided into train, test, and validation sets.

Utility feature was developed to copy the actual files from the source files to the destination directories.

```python
#Utility function to copy images to destination folder
def move_files_to_folder(list_of_files, destination_folder):
    for f in list_of_files:
        try:
            shutil.copy(f, destination_folder)
        except:
            print(f)
            assert False
```

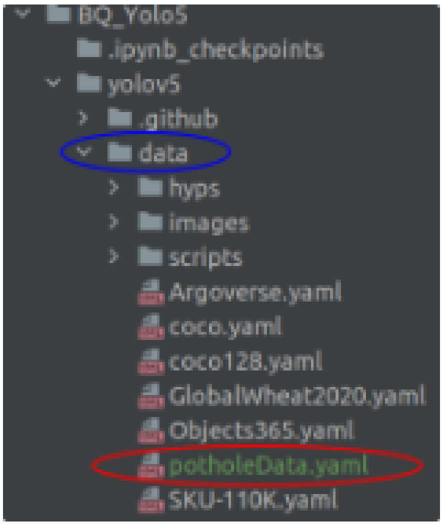
Using the aforementioned utility function, copy the files.

```python
# Copy the splits into the respective folders
move_files_to_folder(train_images, 'potholeData/images/train')
move_files_to_folder(val_images, 'potholeData/images/val/')
move_files_to_folder(test_images, 'potholeData/images/test/')
move_files_to_folder(train_annotations, 'potholeData/labels/train/'
move_files_to_folder(val_annotations, 'potholeData/labels/val/')
move_files_to_folder(test_annotations, 'potholeData/labels/test/')
```

**Training the model.**

A unique file called .yaml file, which contains information on the paths to the train, test, and val files as well as the class names, must be established before the training process can begin.

```yaml
train: /BayesianQuest/Pothole/yolov5/potholeData/images/train/
val:  /BayesianQuest/Pothole/yolov5/potholeData/images/val/
test: /BayesianQuest/Pothole/yolov5/potholeData/images/test/

# number of classes
nc: 4

# class names
names: ["pothole","vegetation", "sign","vehicle"]
```




Start training.

```
!python train.py --img 640 --cfg yolov5m.yaml --hyp data/hyps/hyp.s
```

Output

```
Image sizes 640 train, 640 val
Using 4 dataloader workers
Logging results to runs/train/yolo_pothole_det_m3
Starting training for 500 epochs...

     Epoch   gpu_mem       box       obj       cls    labels  img_size
     0/499        0G    0.1181   0.03669   0.02818        12       640: 100%|
               Class     Images     Labels          P          R     mAP@.5 mAP@WARNING: NMS time limit 0.160s ex
ceeded
               Class     Images     Labels          P          R     mAP@.5 mAP@
                 all          2          0          0          0          0          0

     Epoch   gpu_mem       box       obj       cls    labels  img_size
     1/499        0G    0.1214   0.03416   0.02628         9       640: 100%|
               Class     Images     Labels          P          R     mAP@.5 mAP@WARNING: NMS time limit 0.160s ex
ceeded
               Class     Images     Labels          P          R     mAP@.5 mAP@
                 all          2          0          0          0          0          0
```

The following command should be entered on a terminal to visualize the training, which is a time-consuming operation, on Tensorboard. The runs/train folder contains the log information needed to execute Tensorboard.

```
Pothole/BQ_Yolo5/yolov5$ tensorboard --logdir runs/train
```

```
NOTE: Using experimental fast data loading logic. To disable, pass
    "--load_fast=false" and report issues on GitHub. More details:
    https://github.com/tensorflow/tensorboard/issues/4784

Serving TensorBoard on localhost; to expose to the network, use a proxy or pass --bind_all
TensorBoard 2.5.0 at http://localhost:6006/ (Press CTRL+C to quit)
```

Similar output on the browser.


Samples.




The bounding boxes have successfully localized, as can be seen. It is clear that fewer photos were used, yet nevertheless some decent results were obtained. It could be achieved better outcomes with additional photos.

## 5.2 TESTING

### 5.2.1. Test Plan and test strategy

Test planning is required for creating a baseline plan with tasks and accomplishments to track the project's progress. It's described the test approach, goals, agenda, estimation, deliverables, and assets required to perform trying out for a software product. Furthermore, the functions to be tested are chosen based on the importance of the functions and the risks they pose to users. The test cases were then written under the available use cases. They were handled manually, and the results were recorded.

Making a test plan document has many advantages.

• Help customers, business managers, and other non-test team members understand the specifics of testing.

• Test Plan guides our thinking. It looks like a series of rules that must be adhered to.

• Information on test estimation, test scope, and test technique is included in the test plan. The Management Team can review them and apply them to later projects.

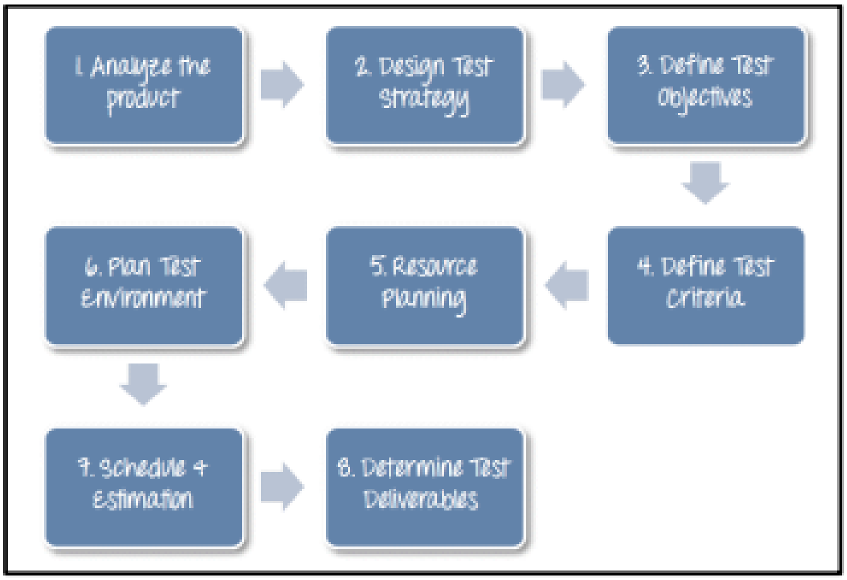



Figure 5. 22: Test planning steps

# 6. REQUIREMENTS

## 6.1. FUNCTIONAL REQUIREMENTS

Table 5.1.1.: functional requirements

| Requirement ID | The Requirement | Addressing the Requirement |
|---|---|---|
| 1 | Extract features | The video will be split into frame and the images will be used to detect the of the objects. |
| 2 | Identify obstacles | Using real time images for detect the obstacle using by DNN model + SSD. |
| 3 | Identify walking surface | Using real time images for detect the walking surface nature using by DNN model + SSD. |
| 4 | Make a warning message for close objects. | According to obstacle distance give user to waring message. |
| 5 | Generate a voice notification about obstacles and surface. | After identifying the obstacle and surface, convert them into voice command. |

## 6.2. NON-FUNCTIONAL REQUIREMENTS

| Requirement ID | The non-functional Requirement |
|---|---|
| 1 | **Security**- This non-functional requirement ensures that all data is protected from ransomware attacks and unauthorized access inside the device or a portion of it. |
| 2 | **Performance** - Defines how quickly a software system or one of its core components responds to the actions of certain users during a given workload. |
| 3 | **Reliability** - This quality feature estimates the likelihood of the device or its component operating without failure under predefined conditions for a specific amount of time. It's usually represented as a percentage of the possibility. For example, if the device has an 85 percent dependability for a month, there is |

| | |
|---|---|
| | an 85 percent chance that the system will not extinguish during that month under normal operating settings. |

Table 5.1.2.: non-functional requirements

**User Requirements**

- User should be able to identify obstacles on their way.
- User should be notifying about the behaviour of the walking surface.
- User should be warned if the objects are too close.
- User should be able to operate the system independently.
- User should be able to get voice assistance.

# 7. RESULT AND DISCUSSIONS

Several practical tests have been conducted in various environments and individuals, and the results have been obtained to evaluate the system's performance. The face recognition program could identify individuals in a close range with high accuracy; however, the accuracy gradually declined as the individual was out of clear range for the camera. Nevertheless, upon increasing the camera's quality, the system could identify an acceptable range for the user. The emotion detection model currently works only near the user as facial features should be visible to identify the correct emotion displayed by the individual. The walk-lane detection system was created to aid those who are blind or visually impaired when moving through the outside world.

The proposed gadget is more convenient for visually challenged people because the microcontroller is attached to the eyeglasses. Users can securely navigate the walk lanes using an integrated voice command feature. With more than 80% accuracy, this system offers a straightforward, practical, and dependable answer to the users' travel-related issues. A road sign detection system has been conducted with an 80% accuracy rate with shortcomings in the inability to identify the road sign correctly due to discoloration and improper alignment of signs. Obstacle detection and pothole detection using a smart cane has been conducted using an alarm notification to identify potential obstacles and potholes on roads, but the team intends to switch to voice notifications for further efficiency in the system.

Though the system is accurate, there are some shortcomings, such as being unable to use it in dim light and the nighttime. Another limitation of the system is that except for the smart cane detection of potholes, the user must be facing a particular direction to identify road signs, individuals, obstacles, or walk lanes.

# 8. DESCRIPTION OF PERSONAL AND FACILITIES

The suggested system will use image processing to model object detection. For blind users, this program will be created in a more user-friendly and effective manner. The application's accuracy and efficiency will be improved as a result of the combination of several methods. This application will be built using open-source computer vision (OpenCV) techniques such as background subtraction, motion detection, edge detection, and feature extraction. The image processing findings will be translated into audio directions for the user when the barriers are spotted using the mobile phone's camera. When compared to comparable equipment on the market, the proposed application will be a low-cost application. As a result, this system will boost blind people's confidence by providing them with accurate information about the items around them and allowing them to walk around independently in both indoor and outdoor environments.

# 9. FUTURE SCOPE

The following are some tips that may be taken into account for upcoming activities involving road damage detection:

The methods of training and the pool of training data currently available place restrictions on the task of detecting road damage. The usage of improved techniques on any additional fresher datasets is possible in the future.

This research mainly focuses on the topic of surface road damage identification and categorization utilizing an Android device.

Adding new features like severity analysis and pixel-level damage analysis for a larger spectrum of surface road problems could pose further difficulties.

To properly analyze the performance, other evaluation measures can be used. For example, for building quicker real-time smartphone-based object recognition, taking into account In addition to F1-Score, other important factors include the trained model's inference rate and disk capacity.

Future iterations may take into account the models' performance for various damage classes in addition to the average F1-score. A more balanced representation of damage classes could be added to the road damage dataset.

- We are hoping to use raspberry pi4 module for the future implementation of the component.
- Tune the model to those data samples and get higher results.
- Integrating all the four component to a single unit.

# 10. CONCLUSION

The Smart Navigation system has been primarily developed using the concept of Image Processing and also uses the concept of IoT for the smart cane implementation. The application has been developed with a user-friendly interface to effectively aid the visually impaired community. Due to multiple features such as face recognition coupled with emotion detection, walk-lane detection, road sign detection, obstacle detection, and pothole detection using the smart cane, the current system presents a unique opportunity for blind people to become more active in a cost-effective manner. The features detected through the camera and its image processing results are converted into audio/vocal commands and relayed to the user. Compared to the other systems in the market, this system consists of several unique features that cater to the multiple needs of blind people and encourage them to lead a more active social life. However, there are several shortcomings in the application. Therefore, overcoming these shortcomings and making this application more usable for the blind community is essential. Road damage identification is a significant issue, and numerous types of study have been conducted to overcome this difficulty. By using data from Sri Lankan roads to train the model, we applied a YOLO-based technique to deep learning to detect road deterioration. Smartphones are used to collect the dataset. We examined numerous dataset settings, and the results revealed some intriguing details. One pre-trained weight only requires 42MB of memory for our YOLOv5-based solution, and its inference performance is very quick. In the context of actual road damage identification issues, not only accuracy but also inference speed and, in some cases, frame rate are key considerations. As a result, this approach might be a good choice for blind people to detect road damage.

## 11. REFERENCES

[1] https://www.iapb.org/learn/vision-atlas/

[2] Blind Navigation System Using Image Processing and Embedded System.

[3] Hersh, M., Johnson, M.A.: Assistive Technology for Visually Impaired and Blind People, 1st edn. Springer, London (2008)

[4]file:///C:/Users/grace's%20secret/Desktop/Research/ResearchPapers/Obstacle%20detection%20and%20warning%20system%20for%20visually%20impaired.pdf

[5] Solomon, N., Bhandari, P.: Paten lanscape report on assistive devices and technologies for visually and hearing impaired persons. Technical report, Patent lanscape report project (2015)

[6] Hollins, M., 1989. Understanding blindness: An integrated approach. Lawrence Erlbaum Associates, Hillsdale, NJ. 1

[7] Manduchi, R., and Kurniawan, S., 2011. “Mobility-related accidents experienced by people with visual impairment.” Insight: Research and Practice in Visual Impairment and Blindness, 4. 1

[11] Manikanta K S, Phani T S S and Pravin A. 2018. Implementation and Design of Smart Blind Stick for Obstacle Detection and Navigation System International Journal of Engineering Science and Computing 8 (8).

[12] Amjed S. Al-Fahoum, Heba B. Al-Hmound and Ausaila A. Al-Fraihat. 2013. A Smart Infrared Microcontroller-Based Blind Guidance System Hindawi Publishing Corporation Active and Passive Components 2.

[13] Ayat Nada, Samia Mashelley, Mahmoud A Fakhr and Ahmed F Seddik. 2015. Effective Fast Response Smart for Blind People Second International conference in Bioinformatics and environmental engineering.

[14] Vijayalakshmi Badre, Romba Chhabria, Tanmay Kandam and Kritika Karamchandani. 2016. Ultrasonic Blind Walking Stick with Voice Playback International Research Journal of Engineering and Technology 03 (04).

[15] Somnath Koley and Ravi Mishra. 2012. Voice Operated Outdoor Navigation System for Visually Impaired Persons International Journal of Engineering Trends and Technology 03 (02).

[16] Yashas, M., Maalik, K.S., M.Z.M., A., & K.P.S.H., D.S. (2017). Blind Guider: An IT Solution for Visually Impaired People. International journal of scientific and research publications, 7.

[17] R. Kedia et al., "MAVI: Mobility Assistant for Visually Impaired with Optional Use of Local and Cloud Resources," 2019 32nd International Conference on VLSI Design and 2019

18th International Conference on Embedded Systems (VLSID), 2019, pp. 227-232, doi: 10.1109/VLSID.2019.00058.

[18] P. Meijer, “An Experimental System for Auditory Image Representations,” IEEE Transactions on Biomedical Engineering, vol. 39, no 2, pp. 112-121, Feb 1991.

[19] G. Sainarayanan, “On Intelligent Image Processing Methodologies Applied to Navigation Assistance for Visually Impaired”, Ph. D. Thesis, University Malaysia Sabah, 2002.

[20] . Balakrishnan, G. Sainarayanan, R. Nagarajan and S. Yaacob, “Wearable Real-Time Stereo Vision for the Visually Impaired,” Engineering Letters, vol. 14, no. 2, 2007.

[21] G. P. Fajarnes, L. Dunai, V. S. Praderas and I. Dunai, “CASBLiP- a new cognitive object detection and orientation system for impaired people,” Proceedings of the 4th International Conference on Cognitive Systems, ETH Zurich, Switzerland, 2010.

[22] Manikanta K S, Phani T S S and Pravin A. 2018. Implementation and Design of Smart Blind Stick for Obstacle Detection and Navigation System International Journal of Engineering Science and Computing 8 (8).

[23] Amjed S. Al-Fahoum, Heba B. Al-Hmound and Ausaila A. Al-Fraihat. 2013. A Smart Infrared Microcontroller-Based Blind Guidance System Hindawi Publishing Corporation Active and Passive Components 2

[24] Ayat Nada, Samia Mashelley, Mahmoud A Fakhr and Ahmed F Seddik. 2015. Effective Fast Response Smart for Blind People Second International conference in Bioinformatics and environmental engineering.

[25] Vijayalakshmi Badre, Romba Chhabria, Tanmay Kandam and Kritika Karamchandani. 2016. Ultrasonic Blind Walking Stick with Voice Playback International Research Journal of Engineering and Technology 03 (04).

[26] Roentgen, U. R., Gelderblom, G. J., Soede, M., and De Witte, L. P., 2008. “Inventory of Electronic Mobility Aids for Persons with Visual Impairments : A Literature Review.” Journal of Visual Impairment & Blindness(November), pp. 702–724. 4, 5, 38

[27] U. Michelucci, "A Case-Based Approach to Understanding Deep Neural Networks," in Applied Deep Learning, Switzerland, Apress, Berkeley, CA, 2018, p. 410.

[28] I. C. Education, "Machine Learning," IBM, 15 07 2020. [Online]. Available: https://www.ibm.com/cloud/learn/machine-learning.

[29] E. Burns, "In-depth guide to machine learning in the enterprise," Tech Accelerator, 2018.

[30] I. CloudBolt Software, "What Is A Cloud Platform?," CloudBolt, 2021. [Online].

[31] Object (image processing) [Online], Available at: https://en.wikipedia.org/wiki/Object_(image_processing) , [Assessed: 02/10/20]