Machine learning for visually impaired people using object detection

By:- Ayush Kumar

30/08/2022

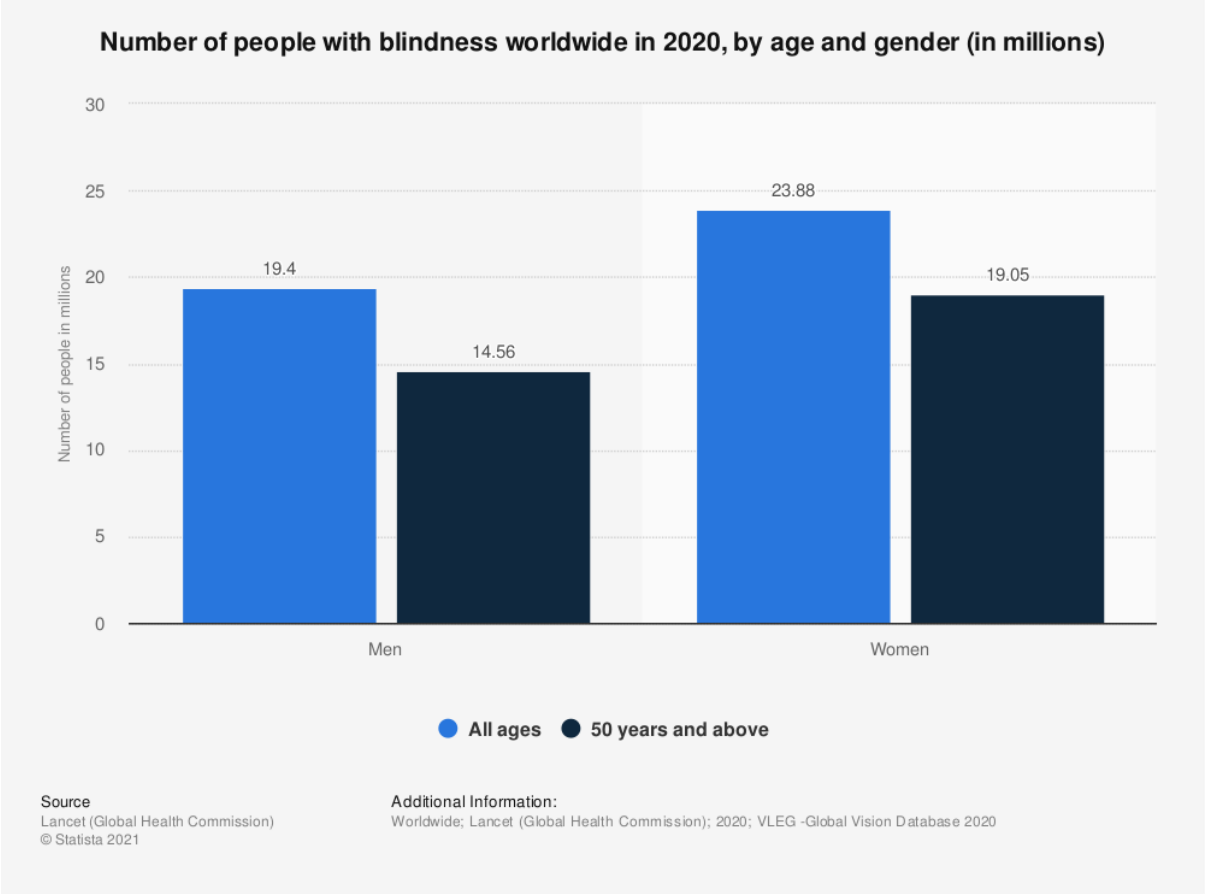
Abstract

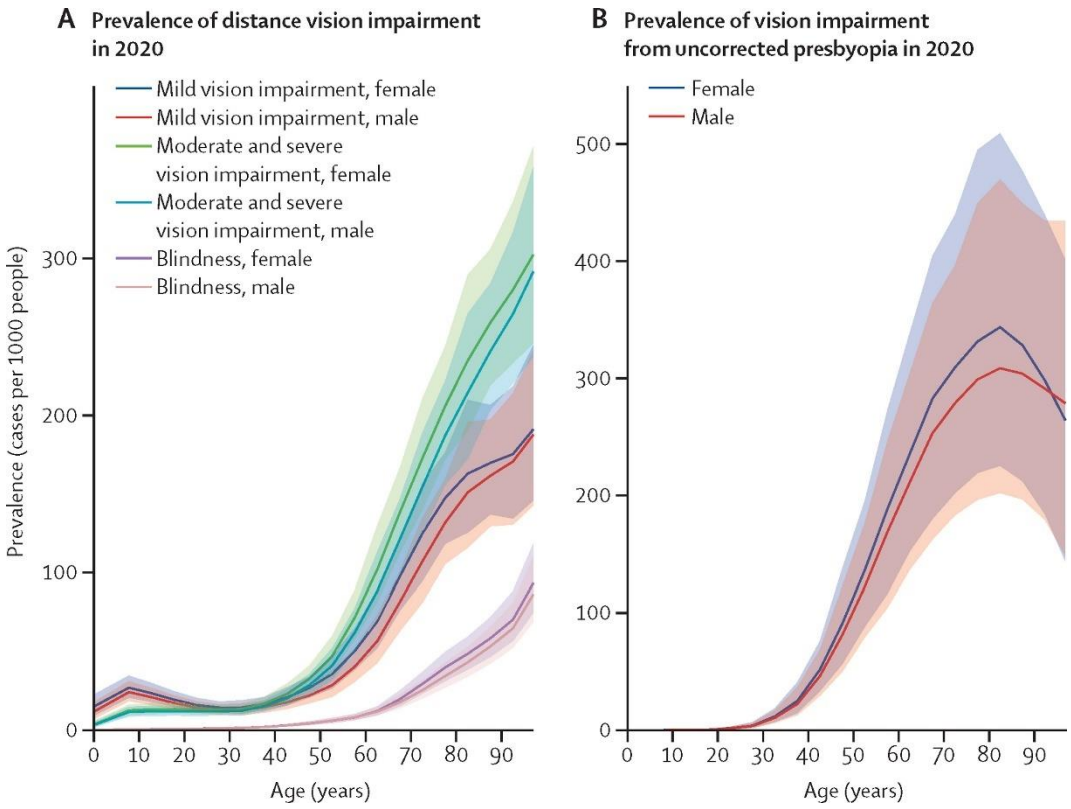
In this report, I have proposed the idea of using object detection, there have been many areas where technology has proved itself as a special need for humans. From starting the day with Alexa to ending it with Alexa, we all are so involved in technology. There are almost 39 million people who don't have the blessing of seeing and watching things like normal people but due to machine learning, it is possible that even they can also detect things not directly but indirectly using machine learning and artificial intelligence.

Degradation of the visual system can lead to a dramatic reduction of mobility by limiting a person to his sense of touch and hearing. This report presents the development of an obstacle detection system for visually impaired people. While moving in his environment the user is alerted to close obstacles in range. The system detects an obstacle surrounding the user by using a multi-sonar system and sending appropriate vibrotactile feedback. The system aims at increasing the mobility of visually impaired people by offering new sensing abilities.

1. Problem Statement

The problem statement is to apply machine learning to help blind people to recognize objects and help them to travel without others' help. Moving into an unknown environment becomes a real challenge when we cannot rely on our own eyes. Since dynamic obstacles usually produce noise while moving, blind people develop their sense of hearing to localize these. However, they are reduced to their sense of touch when it is a question of determining the exact position of an inanimate object. The common way of navigating for a visionless person is to use a white cane or a walking cane. The walking cane is a simple and purely mechanical device dedicated to detecting static obstacles on the ground, uneven surfaces, holes, and steps via simple tactile-force feedback.





2. Assessment

2.1) Customer need /people need assessment

There are many devices available in the market to support a good vision for visually impaired people but either they are too expensive or they are not available in adequate quantity. These kinds of devices also do not generate good output on which the customer can trust upon. The need of the hour is to invent new technology and device which is much more affordable and has much more accuracy than older devices. This can be possible with the advancement of machine learning, deep learning, and artificial intelligence. There are many challenges faced by blind people like

the lack of emotional support at diagnosis centers, the limited accessibility to activities and information, the societal stigma, and the lack of unemployment, which are all factors that frequently lead blind or low-vision individuals in isolation.

We can use object detection techniques to help visually impaired people to recognize the obstacles on the road. Therefore, by using this technique, we aim to provide a robust solution for visually impaired people using which they will be much more independent and will have access to much more opportunities from which they are still unfamiliar.

2.2) Market need assessment

Assistive Technologies for Visually Impaired Market size was valued at USD 799.21 Million in 2020 and is projected to reach USD 1,364.57 Million by 2028, growing at a CAGR of 6.92% from 2021 to 2028. The growing incidences of visual impairments and blindness, an aging population, rising awareness of assistive technologies for the visually challenged, and technical developments are driving the Global Assistive Technologies for the Visually Impaired market forward during the predicted period. In addition, increasing health care expenses, along with increased funding and growing government initiatives to help encourage good facilities, are anticipated to propel the global Assistive Technologies for Visually Impaired Market growth further over time. The Global Assistive Technologies for Visually Impaired Market report provides a holistic evaluation of the market. The report offers a comprehensive analysis of key segments, trends, drivers, restraints, competitive landscape, and factors that are playing a substantial role in the market. Assistive technology is referred to as a

device that enables disabled or visually impaired individuals to overcome physical limitations they may be suffering from to accomplish a task. As compared to different supporting components of a computer, there is a large variety of alternative or modified keyboards, screen magnifying software, and text-to-speech software that provides greater access to the visually impaired.

2.3) BUSINESS NEED ASSESSMENT

In this assessment, we find out what visually impaired people needs. It defines the gaps that are preventing the technology from reaching its desired goals. It also contains the strategy to make this business perfect or up to the mark using different optimization processes.

3. Target Specifications

In this paper, we aim to target the outcomes of using technology for inventing new ideas and concepts for helping and supporting visually impaired people. Object detection is a concept in deep learning which uses different machine learning algorithms for implementing and detecting the different objects in a given image. We will use the YOU ONLY LOOK ONCE state-of-the-art algorithm for building the boundaries which will define the objects in the given image, it makes a rectangular or square box around the image, and the Yolo algorithm classifies it.

1. Using machine learning and different deep learning algorithms for implementing object detection -
   Image classification involves assigning a class label to an image, whereas object localization involves drawing a bounding box around

one or more objects in an image. Object detection is more challenging and combines these two tasks and draws a bounding box around each object of interest in the image and assigns them a class label. Together, all of these problems are referred to as object recognition.

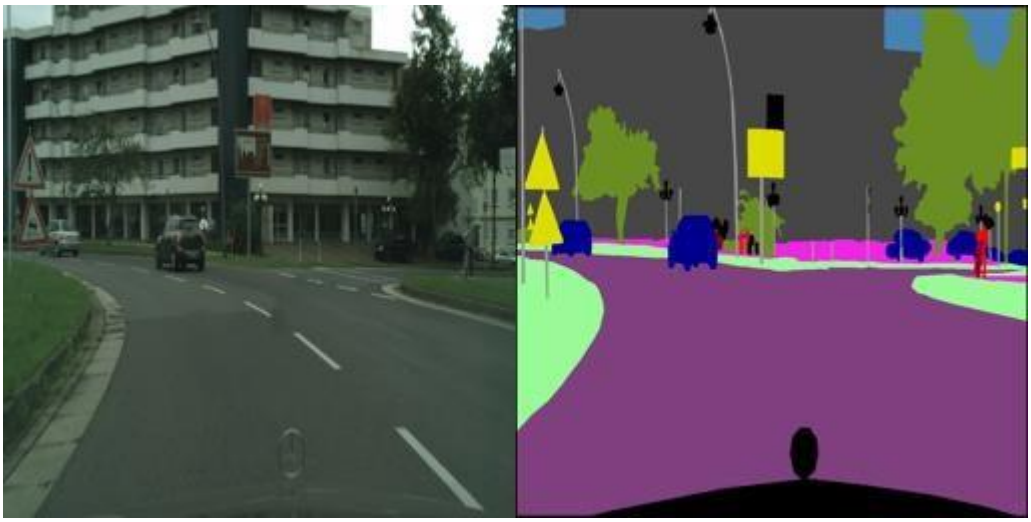


Fig showing the basic idea.

2. Developing a wearable device that can be used to achieve the goal.

   The system consists in sensing the surrounding environment via sonar sensors and sending vibrotactile feedback to the user of the position of the closest obstacles in range. The idea is to extend the senses of the user through a cyborgian interface. This means that the user should use it, after a training period, without any conscious effort, as an extension of his own body functions. Since the visually impaired community is reluctant toward new technologies designing this concept would be a complement to the traditional white cane. It will focus on detecting obstacles at shoulder height and on letting the user have his hands free.

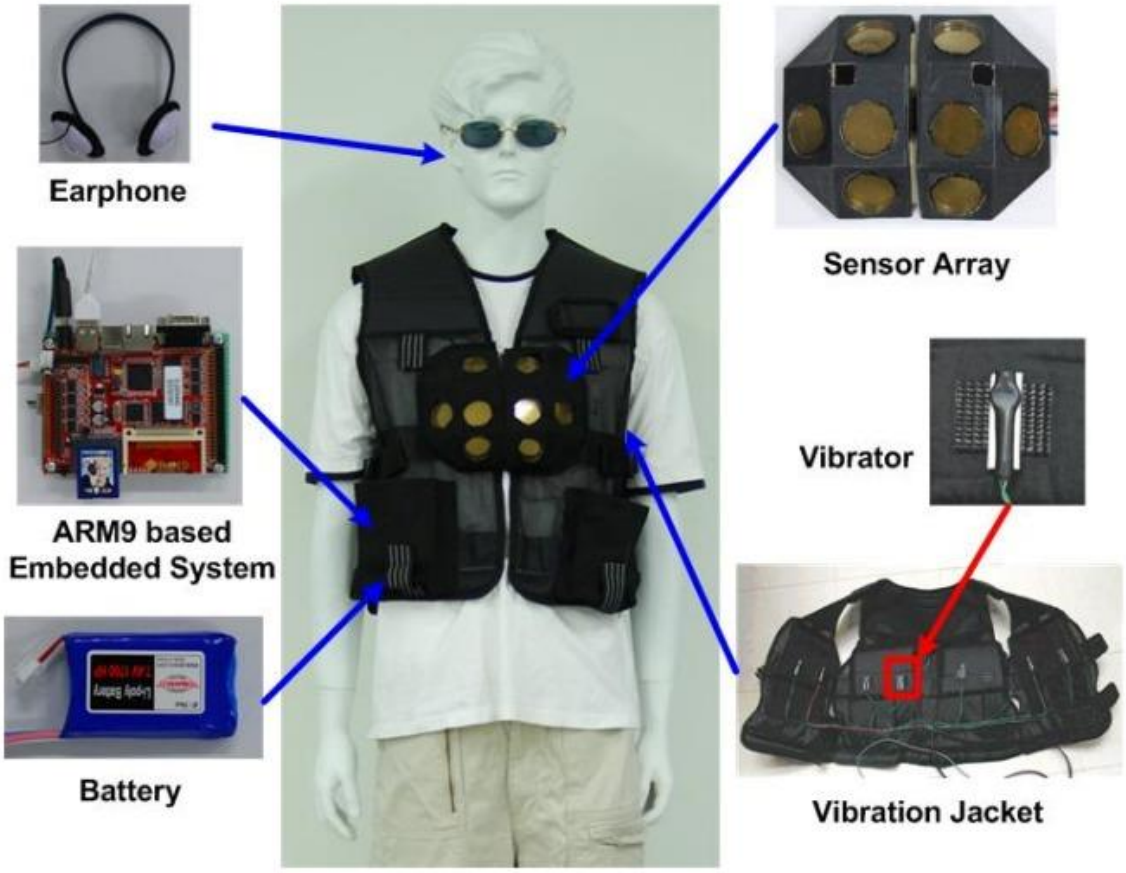


Fig depicting the basic elements of the device.

3. <u>The predictive model predicts the outcome of the given image.</u>

   To predict the outcome of a given image we can use some algorithms of machine learning like Convolutional Neural Networks (CNN) which are specially designed for this classification purpose. We can also use YOLO (You only look once) which is a real-time object detection algorithm to differentiate between the different objects in the given image. Some machine learning frameworks which can be used are TensorFlow, Keras, etc.

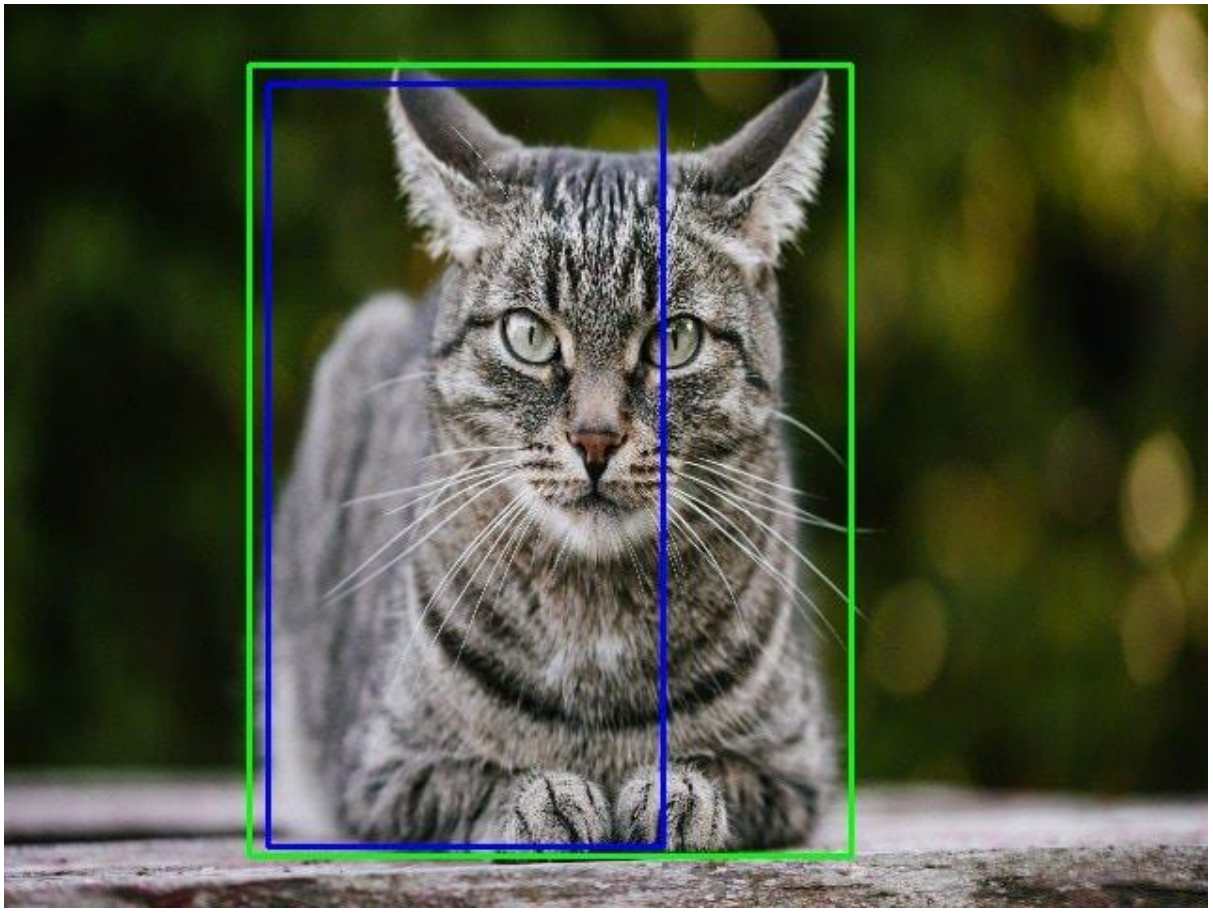


Fig showing some principles of YOLO

4. Using natural language processing (NLP) to understand human language can further be translated to machine translation which will help visually impaired people to understand the signs and commands.

   Image analysis is the extraction of meaningful information mainly from digitally stored images through several processing techniques, however, may be useful for a small range of tasks as there still aren't any known methods of image analysis that are generic enough for a wide range of tasks, compared to the abilities of a human's image analyzing capabilities. Natural language processing (NLP) on the other hand is a part of artificial intelligence where we apply computational techniques to the analysis and synthesis of natural language and speech. NLP tasks can be separated into low-level tasks and high-level tasks. Some of these tasks have direct applications, while others are sub-tasks that are used to help solve larger tasks, with low-level tasks usually feeding into high-level tasks.

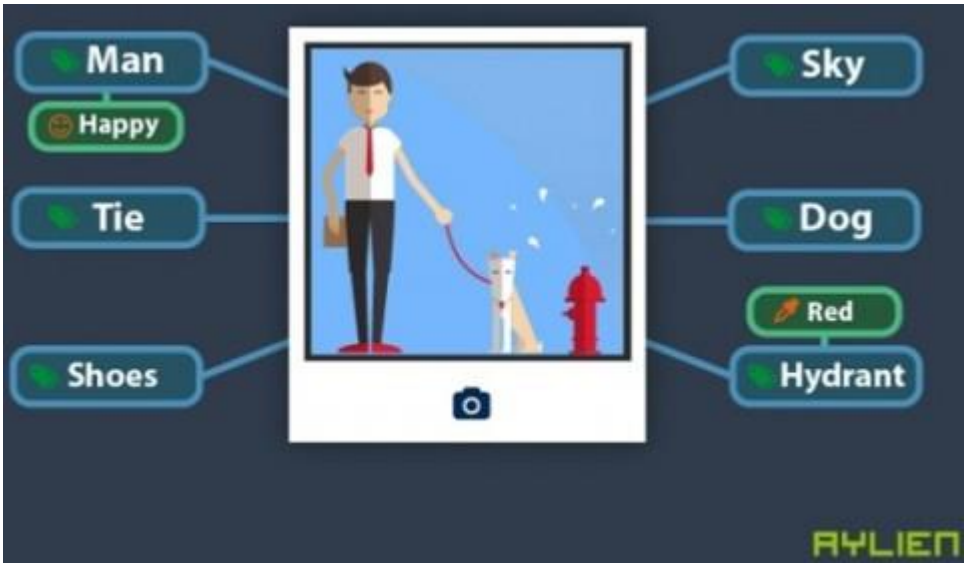


Fig showing the basic idea of NLP and image analysis

5. External Search

The external sources which I have used as a reference for understanding and implementing different methods for obstacle detection for visually impaired people and how recent research are going on to achieve this goal are as follows

1. How machine learning can be used for blind people.
2. Wearable Obstacle Detection System for visually impaired People
3. Challenges faced by blind people when living life.
4. Studies on the number of blind people in the world.
5. AI-based vision research on blind people.
6. Assistive technologies for visually Impaired market size and forecast.

6.1 Benchmarking

Some of the companies which are involved in making devices using different technologies like TCS(Tata Consultancy Services), Amway, Fullerton India, and Galalite make use of these technologies for developing different devices which can also be used by small companies or start-ups to get an edge over bigger companies by working on the elementary grass root level and can bring a huge change by developing affordable and more trustworthy devices.

6.2 Applicable patents

Some of the patents which can be applicable here whose goals overlap with our goals can be as follows –

Intelligent glasses for visually impaired people.

An approach for communicating navigation information on a physical environment to a user. The approach includes a computer receiving digital images of the physical environment of the user captured by digital video devices and converting the digital images into a three-dimensional image. The approach includes the computer analyzing the three-dimensional image using objective analysis to generate output data, wherein the output data corresponds to the physical environment and determining at least one device associated with the user. The approach includes the computer formatting the output data for use with at least one device, wherein the at least one device is capable of providing to the user a spatial map of the physical environment created by an electrical stimulation pad, and receiving a touch from a user on a surface of the electrical stimulation pad corresponding to an object represented in the spatial map of the physical environment.

Obstacle Avoidance System for Visually Impaired People.

According to the World Health Organization(2014), 285 million people are estimated to be visually impaired worldwide: 39 million are blind and 246 million have low vision. Electronic travel aids (ETAs) have the potential to increase the mobility and with that the quality of life of the visually impaired. ETAs have two main components: (1) sensors to determine location and

orientation and to detect and possibly identify objects in the environment and (2) displays to present information about for instance waypoints and obstacles. Since visual displays are of little or no use to this user population, work has been focused on displaying information through the sense of hearing and/or touch. In this paper, we focus on the design of a multisensory display for an ETA and not on its sensors.

There are many more such papers that focus on the technology to be used to implement the given idea, but I found these two papers much more informative and more relevant as compared to other papers.

6.3 Applicable Constraints

- Lack of proper datasets.
- Don't depend on the colors of the input data itself to detect the object.
- Sufficient contrast and color of the image.
- A balanced detailed image is really hard to find.
- Special attention to less recognized signs or objects.
- The model should be able to work in even fewer lights and should be able to work in fog during winters.
- Requires a very fast working model to recognize incoming objects or obstacles.
- Convincing other companies to consider the technology and implement it.

- A lot of governmental paperwork is to be done and authorizations to be taken as it is a very sensitive device.

6.4 Applicable Regulations

- Selection of sources of legislation and trends.
- Selection of relevant legislation.
- Data protection of the user and the consumer.
- Keeping the track of taxation and accounting laws.
- Taking care of the contracts which the company is going to sign with other organizations.

7. Business Opportunity

These technology ideas can be implemented by small companies or start-ups as they are relatively new and has much more demand as compared to other technologies or devices of this segment. This is a social issue that requires support and if given proper guidance and effort this can turn out to be a good business idea as this technology has fewer competitors in the market and this opportunity along with a good well proper planned strategy can be used to establish a good reputation in the market of visually impaired devices.

8. Final product prototype

The final product would be a machine learning model or technology which will help the device recognize the different obstacles on the

way, and this will ultimately lead to a completely new defined market size for visually impaired people and other companies who are involved in making these kinds of devices

The service implements a machine learning model using object detection using the YOLO algorithm for object detection in a given image. The object recognition system can be realized by semantic segmentation, that is, to assign semantic labels (such as a tree, or a sidewalk) to every pixel in the input image.

1. The object detection system is a part of deep learning.
2. The preparation of the dataset is an essential part of the model implementation.
3. The selection of a well-optimized machine learning algorithm is another important step for model implementation.
4. Uses a smartphone for the software implementation and uses it to capture live images of the surroundings and send them as input to the model.
5. It provides voice support and guidance to help visually impaired people to recognize the existence of obstacles on the path in real-time.

Implementation of the above approach

1. Selecting the dataset which contains images of any city.
2. Cleaning the dataset and separating the training and testing data for the model implementation.
3. It is possible that there can be more than 10k different colors in a single image, so we will need to normalize the colors to the nearest color to which it might be similar.
4. After the pre-processing of the training data, now it is time to select the model architecture for its implementation.

5. For training purposes, we will build a semantic segmentation model using deeplabv3+ on the given dataset.
6. We can use this model to recognize objects in the given image.
7. After passing the input through the model, now our model is ready to recognize the image and can give the expected output.

9. Conclusion

So, concluding this report, this report answers the question that how can we use machine learning, deep learning, and artificial intelligence for helping visually impaired people. We can use object detection and can apply this model to the input image given by the device or the user then this model will give instructions to the user according to the scenario of the output image which will come after processing through the model.

I have hence proposed the application of this technique for small companies and start-up companies. This is not a fully robust, checked plan but with some considerable efforts and techniques, this can set a milestone in the industry of devices for visually impaired people.