

Model for assisting blind people by detecting the Surrounding Objects / People/ Activities with voice description

Raahul Varman¹, Suhas Preetham Kambham²

^{1,2}Department Of Computer Science, Amrita University, Coimbatore, Tamil Nadu, India

[1cb.en.p2cse21012@cb.students.amrita.edu](mailto:cb.en.p2cse21012@cb.students.amrita.edu) [2cb.en.p2cse21008@cb.students.amrita.edu](mailto:cb.en.p2cse21008@cb.students.amrita.edu)

Abstract— Numerous methods have been created to help visually impaired persons and enhance their quality of life. Unfortunately, the majority of these systems have constrained limitations. We made the decision to create a model that can identify any conceivable surrounding objects and direct users through voice description. Our major goal is to identify potential barriers in the person's route and alert them with voice description.

Keywords— Computer Vision, Image Processing, Blind Navigation, Image pixels, CNN, RNN.

I. INTRODUCTION:

The design that can help blind persons navigate the surrounding obstacles in their route. The model's primary goal is to use speech description to identify surrounding impediments in a person's route, such as automobiles, animals, people, and objects, and to lead or warn them.

- **Algorithms:**

Mobile Net – MobileNet is an effective model for mobile and embedded vision applications that builds lightweight deep convolution neural networks using depth wise separable convolutions.

Software/Libraries/Framework:

1. IDE	Jupyter Notebook (Anaconda Navigator) And Google Colab.
2. Library	OpenCV
3. Framework	Tensorflow

Dataset Description:

The MS COCO (Microsoft Common Objects in Context) collection contains extensive datasets for object recognition, segmentation, key-point detection, and captioning. The dataset consists of 328K images broken down into 80 categories. 2014 saw the debut publication of the MS COCO dataset. Totalling 164K pictures, there are 83K for training, 41K for validation, and 41K for testing. The 81K image additional test set that was released in 2015 had all of the earlier test pictures in addition to 40K new ones. Based on feedback, the training/validation ratio was changed from 83K/41K to 118K/5K in 2017. The new split makes use of the same images and annotations. The 2017 test set is made up of a subset of the 41K photographs from the 2015 test set.

person	fire hydrant	elephant	skis	wine glass	broccoli	dining table	toaster
bicycle	stop sign	bear	snowboard	cup	carrot	toilet	sink
car	parking meter	zebra	sports ball	fork	hot dog	tv	refrigerator
motorcycle	bench	giraffe	kite	knife	pizza	laptop	book
airplane	bird	backpack	baseball bat	spoon	donut	mouse	clock
bus	cat	umbrella	baseball glove	bowl	cake	remote	vase
train	dog	handbag	skateboard	banana	chair	keyboard	scissors
truck	horse	tie	surfboard	apple	couch	cell phone	teddy bear
boat	sheep	suitcase	tennis racket	sandwich	potted plant	microwave	hair drier
traffic light	cow	frisbee	bottle	orange	bed	oven	toothbrush

Video Dataset:

100,000 videos are included in the BDD100k Dataset. Each 720p, 30 fps movie lasts for around 40 seconds.

	KITTI	Cityscapes	ApolloScape	Mapillary	BDD100K
# Sequences	22	~50	4	N/A	100,000
# Images	14,999	5000 (+2000)	143,906	25,000	120,000,000
Multiple Cities	No	Yes	No	Yes	Yes
Multiple Weathers	No	No	No	Yes	Yes
Multiple Times of Day	No	No	No	Yes	Yes
Multiple Scene types	Yes	No	No	Yes	Yes

II. Literature Review:

S.no	Title	Abstract	Conclusion
1.	Visual Assistance for Blind using Image Processing [1].	Visually impaired people have a very difficult time navigating daily life. They regularly ask for help from others. For the benefit of persons who are blind, many technologies have been developed. Computer vision-based problem solving are becoming one of the most appealing options among the various technologies being utilised to assist the blind due to its affordability and accessibility. This essay offers a guide for blind people. The proposed system aims to create a wearable visual aid for blind or visually impaired people that can recognise user vocal commands. Its capabilities include item identification and sign board. The blind person will be able to do everyday duties and move around his or her environment more easily as a result. Artificial vision is built on a Raspberry Pi for the Open CV platform using the Python programming language..	This study describes an innovative approach to aiding blind people. The subject can be autonomous in his or her own home thanks to the recommended system's simple architecture and ease of usage. The device also aims to help the blind navigate their surroundings by assisting them in spotting hazards, locating necessities, and reading messages and signs. Initial testing have shown excellent results, enabling the user to move around his environment safely and freely. The technology is made significantly more user-friendly by enabling voice input to access his basic needs.
2.	Assistant for Visually Impaired using Computer Vision [2]	In order to give the user a narrative description of the environment, a virtual assistant for the blind is created in this article employing a variety of technologies, including text-to-speech, object identification, emotion detection, and computer vision. By translating the scenes generated from the visual data shown to the viewer into language that communicates the crucial elements of the scene in audio format, we create an audio story. Future experiments that might use a variety of sensors to do barcode scanning and SLAM-based navigation.	As a result, the finished product is a self-sufficient device that includes a camera, earbuds, and a rechargeable power source. The user's end receives the output as soon as possible. This gadget may be used for a variety of things, including determining the colour of a signal to determine whether it is safe to cross or not, determining the value of cash that is being distributed, and reading road signs directing in the direction of a place the user wants to travel. Therefore, the purpose of this technology is to let the vision handicapped appreciate the lovely world around them while being a bit more autonomous than they were

			before.
3.	Assistive Technology for the Visually Impaired Using Computer Vision [3]	<p>More than a quarter of the 36 million blind individuals worldwide reside in India. One of the most difficult issues facing blind schools today is educating the blind to prevent unemployment among their population. Braille is used in many schools to combat illiteracy, but because of its challenging learning curve, limited availability, and expensive cost, it is mostly out of reach for most students. Less than 10% of India's 12 million blind citizens, according to statistics on braille literacy, are educated in the language. It is clear that the inability to read and learn without the aid of Braille is one of the most important problems that blind and visually impaired people encounter. A system that can help the visually impaired read has to be developed in order to solve this problem. The suggested remedy is to create a low-cost wearable gadget that employs computer vision to read aloud any type of text that is present around the user in a variety of alignments and lighting situations. The device uses a Raspberry Pi and a suitable camera to record the information surrounding the blind or visually impaired person and read it to them in their native tongue. The gadget lists all the items it can see and includes a sensor that notifies the user of the distance to the nearest object at eye level. The system is built using a combination of voice synthesis, machine learning, and image processing approaches. The observed accuracy was determined to be 84 percent when the optical character recognition and the object recognition methods were combined.</p>	<p>The suggested product is effective at capturing readable content in front of the user, identifying the text in the image, and reading it out. The user is also given information about the items around him and the distance of objects that are inside the range of his vision. As a result, this solution facilitates the user's acquisition of information from readable content. He gains the knowledge he needs to be autonomous and to understand his surroundings. The convenient wearable gadget is small and portable. Facial Expression Recognition, Speech Interaction, and Human Face Recollection are a few extensions that may be applied to this product. Consequently, this device might help visually impaired people with daily tasks, improving their lifestyle and making it more like a person without vision loss. Speech interaction, facial expression recognition, and human face recall are a few enhancements that may be applied to this product. Therefore, this device might help visually impaired people with daily tasks, improving their quality of life and making it more like that of someone with normal vision.</p>
4.	Optical Flow in the Dark [4]	<p>The present benchmark datasets for optical flow estimation do not contain enough low-light samples, despite the fact that several efficient optical flow estimation approaches have been developed. These techniques do not perform well when evaluated in low-light situations. The findings for optical flow are still mediocre or even worse even when dark photographs are boosted using preprocessing, which enhances visual perception. This is because information like motion consistency may be lost during enhancement. We offer a comprehensive data-driven method that eliminates error accumulation and quickly deduces optical flow from disorganised low-light images. We build a method designed for simulating the noise model on dark raw images and producing substantial low-light optical flow datasets. We also collect a brand-new optical flow dataset in raw format with a variety of exposure levels to act as a baseline. In comparison to other methods, the models created using our synthetic dataset perform substantially better on low-light images and can mostly maintain optical flow accuracy as image brightness falls.</p>	<p>This article describes our data-driven strategy to improving optical flow accuracy, especially in low-light conditions. On the real low-light optical flow dataset we collect, we successfully train optical flow models that outperform the state-of-the-art. This is accomplished by combining training data from raw images taken at various brightness levels and depending on the noise model we looked at. The purpose of VBOF is to assess the brightness robustness of optical flow models. It consists of 598 raw photographs and the corresponding reference optical flows, each of which has a different brightness. According to us, the noise analysis, the VBOF dataset, and the proposed method will all be very useful for optical flow problems in real environments.</p>
5.	Improved optical flow algorithm of moving object detection [5]	<p>Finding moving targets is one of the most active areas in computer vision research. A crucial factor in determining how successful vehicle collision avoidance warning systems are is the precise recognition of moving objects. A common method for detecting moving targets is an optical flow algorithm. The basic concept and equation regulating optical flow are first covered in this article before a brief explanation of the improved optical flow approach for moving target recognition is given. This work also introduces the double three image interpolation approach and the iterative reweighted least squares method. The results of the experiments show that this method provides more accurate and effective real-time collision detection..</p>	<p>In this paper, they present a more effective algorithm for identifying moving objects and present the calculation formula, general calculation principle, and relevant Matlab programme verification. Matlab simulation results suggest that the algorithm can recognise moving targets more precisely and can roughly estimate its shape when compared to the traditional optical flow approach used in this study. The results of the experiment also show that the target detection performance in real-time is poor, thus the next step should be to speed up detection.</p>
6.	Optical Flow based Obstacle Avoidance for the Visually Impaired [6]	<p>Vision plays a significant role in human navigation. Therefore, regular travel poses a number of challenges for those who are visually impaired. Recognizing and eliminating environmental barriers is the most crucial of them. A broad variety of electronic navigational aids have been created during the past few decades using various technologies for detecting obstacles, such as sonar, infrared, and stereo vision. They have not, however, used optical flow estimations-based navigation, which is being studied in the robotics field and is frequently used by insects. This experiment aimed to evaluate the viability of optical flow estimation-based techniques for guiding a person with vision impairments around barriers using auditory and tactile information. Vision plays a significant role in human navigation. Therefore, regular travel poses a</p>	<p>The calibre of optical ow estimation methods has significantly improved during the past 20 years. Since technology is advancing quickly and our knowledge of nocturnal insect behaviour is improving, it is fair to predict that better methods for estimating motion flow in real time will emerge sooner and help the blind and visually impaired avoid barriers. The purpose of this research has been to evaluate the feasibility of using optical flow estimation methods to develop an auditory and tactile feedback system. The technology has</p>

		number of challenges for those who are visually impaired. Recognizing and eliminating environmental barriers is the most crucial of them. A broad variety of electronic navigational aids have been created during the past few decades using various technologies for detecting obstacles, such as sonar, infrared, and stereo vision. They have not, however, used optical flow estimations-based navigation, which is being studied in the robotics field and is frequently used by insects. This experiment aimed to evaluate the viability of optical flow estimation-based techniques for guiding a person with vision impairments around barriers using auditory and tactile information.	the potential to be used in real-world applications in the future, it was found. A future development might involve looking at the possibility of adopting wearable device manufacturing techniques to aid blind people in getting around.
7.	Real-Time Deep Learning-Based Object Detection Framework	Recently, processing visual input and real-time object detection and identification have become major difficulties in computer vision. Object identification and tracking have been accomplished using a number of approaches in many different industries. On the other hand, conventional classifiers usually deal with challenging tasks where the visual frames are distorted by overlapping, blur from camera motions, changing subject appearances, and ambient factors. Models using OpenCV-based HAAR feature-based cascade classifiers were unable to successfully recognise and track an item in a changing environment because they lacked any error-reducing object identification techniques. Therefore, building a solid integrated framework for real-time object detection and recognition for application in many businesses in the future becomes increasingly crucial. This paper suggests a strong approach for a real-time detector that incorporates Deep Learning Neural Networks in order to reach the maximum degree of computational accuracy (DNN). By eliminating the aforementioned reasons of distortion, the adoption of such a framework will ensure the detector's adaptability and dependability. The model relies on integrating the You Only Look Once (YOLO-v3) object recognition algorithm with the ImageAI deep learning tools and the DarkNet 53 architecture. To ensure dependable data processing, the algorithm was trained using the TensorFlow framework. This research focuses on one crucial component of our long-term goal of creating a multi-agent system since the recommended model will be used in autonomous agents for the detection of landmines, marine debris, and animals in addition to environmental scanning tasks. In this study, tennis balls have been spotted and gathered as a preliminary test for real-world applications to assess the model's effectiveness. The goal of exceeding the accuracy of conventional detectors was nearly achieved by the model.	Recently, academics across many different fields have shown a lot of interest in deep learning-based object recognition systems. Even though we were able to enhance the algorithms related to object detection jobs, there is still room for improvement in terms of accurately recognising multiple and single items. As a result, this paper proposes a proposed model based on the YOLOv3-ResNet detection module of the ImageAI deep learning package. An algorithm that effectively extracts features was successfully developed by combining Darknet-53 with ResNet architecture. According to the testing results, our model performed 68 percent more accurately than conventional object detection methods.
8.	Real Time Object Detection and Tracking Using Deep Learning and OpenCV	Since a few years ago, deep learning has greatly influenced how the world is adjusting to artificial intelligence. Single Shot Detector (SSD), FasterRCNN, Region-based Convolutional Neural Networks (RCNN), and You Only Look Once are a few of the well-liked object identification techniques (YOLO). Among them, Faster-RCNN and SSD perform better in terms of accuracy, but YOLO does better when speed is prioritised above accuracy. For effective application of detection and tracking, deep learning blends SSD and Mobile Nets. Without sacrificing efficiency, our algorithm successfully detects objects.	In real-time circumstances, objects are recognised using the SSD method. Additionally, SSD has demonstrated findings with a high degree of confidence. The primary goal of the SSD algorithm is to identify numerous objects in a real-time video stream and to follow them. The trained item produced good detection and tracking results, and this model may be used in other contexts to find, follow, and react to the targeted objects in the video surveillance. This ecosystem analysis in real time, which enables security, order, and utility for any organisation, may produce excellent outcomes. Increasing the scope of the investigation to look for weapons and ammunition to raise the alert in the event of a terrorist attack. The model may be used in CCTVs, drones, and other surveillance equipment to identify assaults in a variety of locations where weapons are strictly prohibited, such as schools, government buildings, and hospitals.

III. PROPOSED SYSTEM

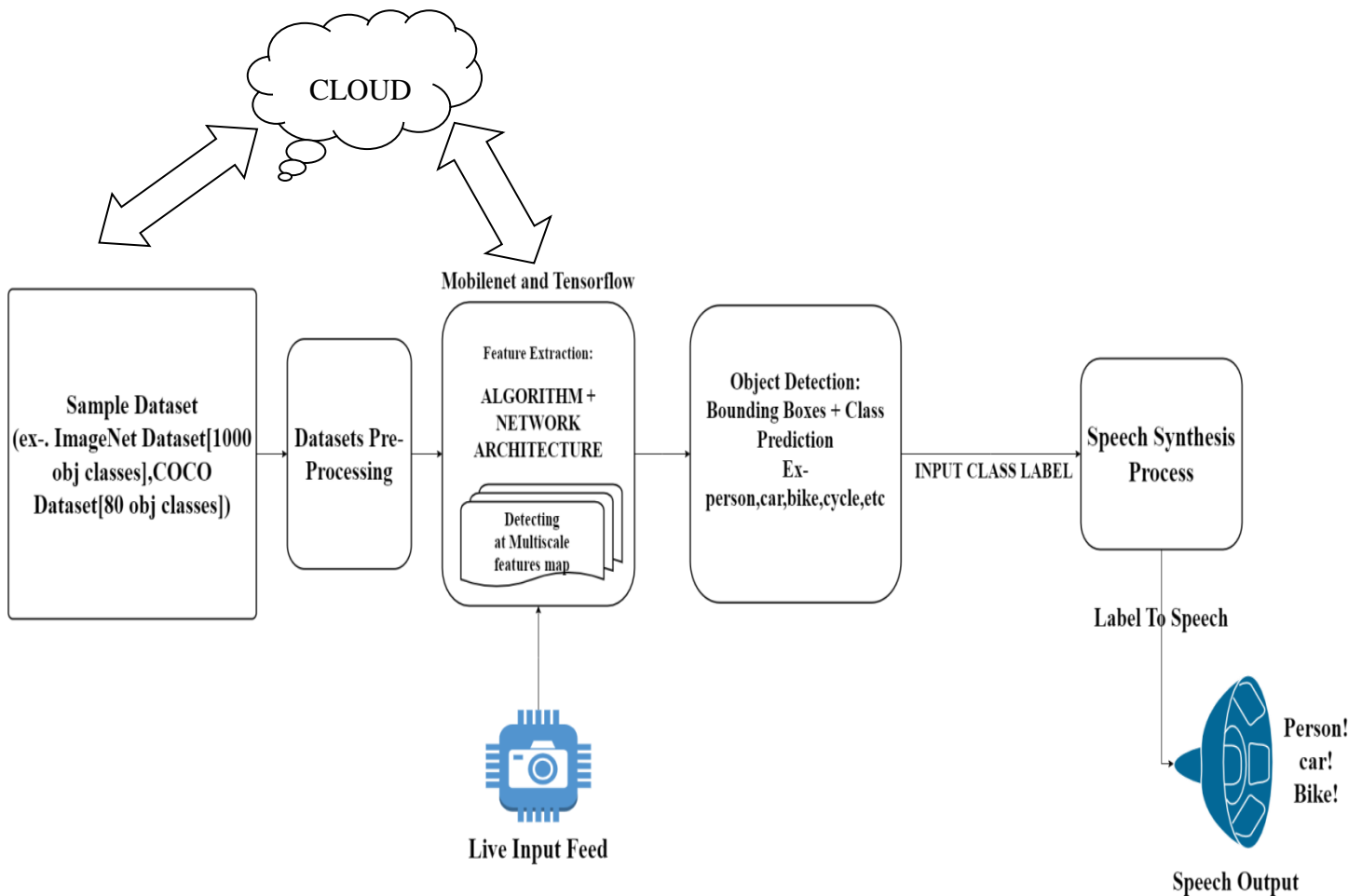


Fig. 1 displays the suggested system's block diagram.

METHODOLOGY :

1. An Model/application will capture real-time videos or/and pictures as part of the system.
2. A pre-trained MobileNet detection model will be used to detect the output class and activity , which was trained on the COCO and BDD100K datasets.
3. Following this , the object's class or the activity recorded will be converted into default voice notes, which will be conveyed to the visually impaired person for assistance.
4. There will be an alarm system that will compute the estimated distance in addition to the object detection. If the visually impaired person is close to the frame or far away at a safer position, it will generate voice-based outputs together with distance units.

MODELS:

The MobileNet model, created for use in mobile apps, is the first mobile computer vision model offered by TensorFlow. Low-latency, low-power models known as MobileNets have been parameterized to accommodate the various use cases' resource constraints. MobileNet makes advantage of depthwise separable convolutions. The number of parameters is considerably decreased when compared to a network with the typical convolutions of the same depth in the nets. Deep neural networks that are lightweight are produced as a

result.

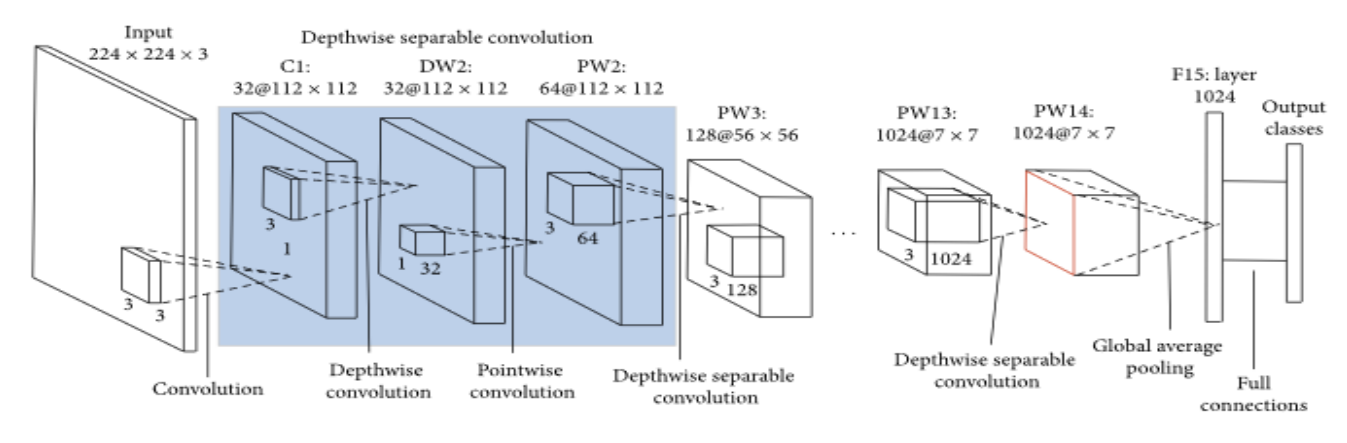


Fig.2.

DEPTHWISE SEPARABLE CONVOLUTION:

The goal behind this convolution was to separate the depth and spatial dimensions of a filter. A depthwise separable convolution is made from two operations : Depthwise convolution and Pointwise convolution.

SPEECH GENERATION MODULE :

Once an object has been found, it is crucial to let the individual know it is there and that they should be on the lookout for it. A key element of the voice generating module is PYTTSX3. Text to voice conversion may be done with Pyttsx3, a Python conversion tool. This method operates as follows: whenever an item is recognised, the texts are shown on the screen. Audio commands are produced as an output.

DATASETS :

IMAGE DATASET : COCO dataset (Common Objects in Context)

Microsoft released a sizable object identification, segmentation, and captioning dataset.

80 object types, or "COCO classes," for which it is simple to identify individual occurrences (person, car, chair, etc.).

91 types of "stuff," where "COCO stuff" refers to substances and things that lack boundaries (such as the sky, the ground, and other objects) but nonetheless convey important contextual information.

VIDEO DATASET : BDD100K Video dataset:

100,000 videos make up the dataset.

Each 720p, 30 fps movie lasts for around 40 seconds.

EVALUATION METRICS :

Performance Metric Name	Formula	Purpose													
Mean Absolute Error (MAE)	<div><div>Divide by the total number of data points</div><div>Predicted output value</div><div>Actual output value</div><div>Sum of</div><div>The absolute value of the residual</div>$MAE = \frac{1}{n} \sum y - \hat{y}$</div>	In regression issues, it is the most basic error metric. The absolute difference between the numbers that were anticipated and those that were actually obtained is basically added up. In other words, we can gauge how inaccurate the forecasts were using MAE. No indication of the model's underperformance or overperformance is provided by MAE, i.e., the model's direction is not shown.													
1) Mean Square Error (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ <div>MSE = mean squared error n = number of data points Y_i = observed values Ŷ_i = predicted values</div>	The sole difference between the MSE and the MAE is that the MSE squares the discrepancy between the expected and actual output values prior to summing them together.													
2) Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (y_j - \check{y}_j)^2}$	One of the methods most frequently used to assess the accuracy of forecasts is root mean square error, also known as root mean square deviation.													
3) Confusion Matrix	<div>Confusion Matrix and ROC Curve</div> <div><table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Predicted Class</th></tr><tr><th>No</th><th>Yes</th></tr><tr><th rowspan="2">Observed Class</th><th>No</th><td>TN</td><td>FP</td></tr><tr><th>Yes</th><td>FN</td><td>TP</td></tr></table><div><div>TN True Negative</div><div>FP False Positive</div><div>FN False Negative</div><div>TP True Positive</div></div></div> <div>Model Performance</div> <div><div>Accuracy = (TN+TP)/(TN+FP+FN+TP)</div><div>Precision = TP/(FP+TP)</div><div>Sensitivity = TP/(TP+FN)</div><div>Specificity = TN/(TN+FP)</div></div>			Predicted Class		No	Yes	Observed Class	No	TN	FP	Yes	FN	TP	When the output of a classification issue can be two or more different types of classes, it is the simplest approach to gauge how well the task is performing. A confusion matrix is nothing more than a table containing two dimensions: "Actual" and "Predicted," as well as "True Positives (TP)", "True Negatives (TN)", "False Positives (FP)", and "False Negatives (FN)" in each of the dimensions.
				Predicted Class											
		No	Yes												
Observed Class	No	TN	FP												
	Yes	FN	TP												

IV. IMPLEMENTATION AND ALGORITHM :

PREPROCESSING TECHNIQUES:

As a part of data preparation, data preprocessing refers to any sort of processing done on raw data to get it ready for another data processing technique. It has long been regarded as a crucial first stage in the data mining process. Data preparation approaches have been modified more recently for the training of AI and machine learning models as well as for making inferences against them.

Data preprocessing alters the data's structure so that data mining, machine learning, and other data science activities can handle it more quickly and efficiently. The methods are typically applied early in the machine learning and AI development pipeline to guarantee correct results.

a. Brightness and Contrast:

Brightness describes the overall lightness or darkness of an image, whereas contrast describes the brightness difference between distinct objects or areas of an image.

The image becomes brighter by adding a positive constant to each of the image's pixel values.

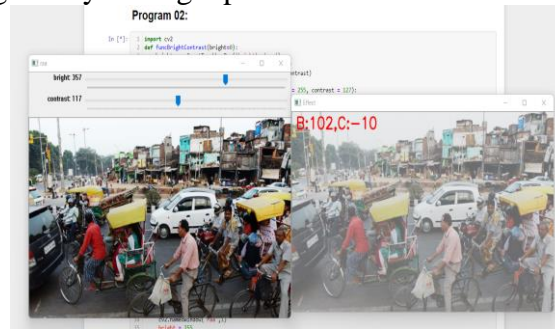


Fig.3. Brightness and Contrast:

b. Histogram Equalization

An essential tool in image processing is a histogram. It is a visual depiction of the distribution of data. The pixel intensity distribution of a digital image is represented graphically by an image histogram. The variable's range of potential values is shown on the x-axis.



Fig.4 . Histogram Equalization

c. Averaging Filter:

By reducing the intensity variation between neighbouring pixels, average filtering is a method for bringing smoothness to images. The average filter goes pixel by pixel through the picture, replacing each value with the average of its neighbours' values, including itself.

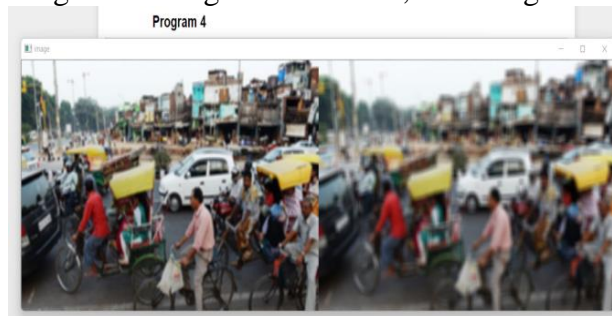


Fig.5. Averaging Filter

d. Median Filter:

A popular non-linear digital filter for reducing visual noise is the median filter.



Fig.6. Median Filter

e. Erosion And Dilation:

Dilation increases the number of pixels around an object's perimeter in a picture, whereas erosion decreases that number. The amount of pixels added or subtracted from the image's objects depends on the size and form of the structuring element used to process it.



Fig.7. Erosion And Dilation

f. Inverting an Image:

In this image processing method, light regions are mapped to dark areas, while dark areas are mapped to light.

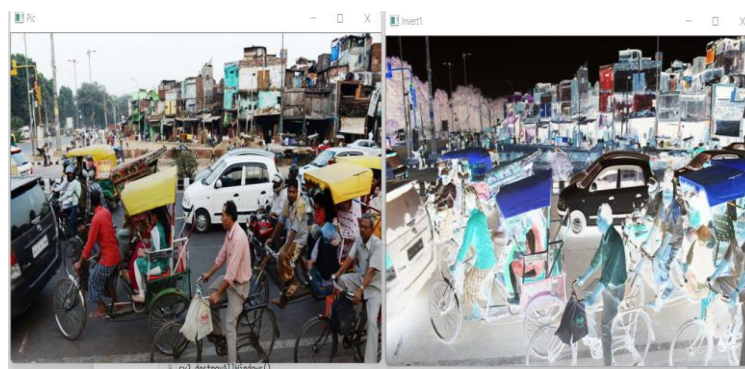


Fig.8. Inverting

FEATURE DETECTION AND FEATURE TRACKING:

No.	Algorithm	Input image	Output image	Principle
1.	Harris Corner Detector			In computer vision techniques, the Harris Corner Detector is a corner detection operator that is frequently used to extract corners and infer features from images.
2.	Scale-Invariant Feature Transform (SIFT)			The local features in a picture, sometimes referred to as the image's "keypoints," may be found with the use of SIFT. These keypoints, which may be utilised for picture matching, object recognition, scene detection, and other computer vision applications, are scale- and rotation-invariant.
3.	Oriented FAST and Rotated BRIEF (ORB)			On the objective of feature detection, ORB outperforms SIFT (and is better than SURF) while being nearly two orders of magnitude quicker. The well-known FAST key-point detector and the BRIEF description serve as the foundation for ORB. Both of these methods are appealing due to their high effectiveness and low expense.

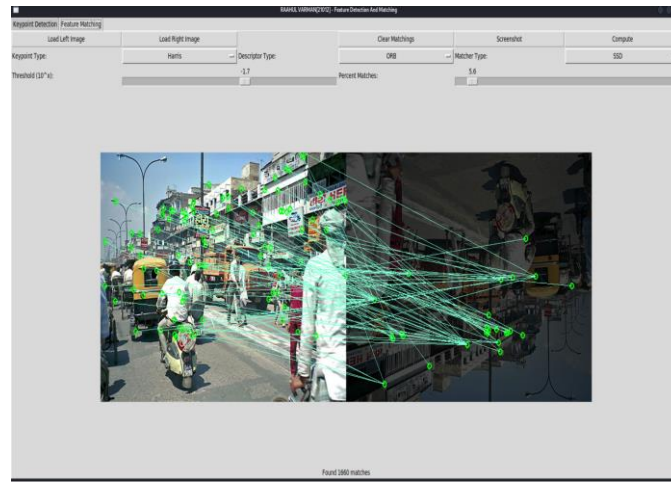


Fig.9 .Feature Matching

a. Metrics used for Feature Matching:

Brute Force Matcher:

We can identify image traits and identifying features in our photos. The second stage is to assign a description from each feature's neighbourhood. Finally, we use descriptors to compare features between two or more photographs. Following that, a variety of tasks, including state prediction, visual odometer, and item recognition, may be performed using the matching attributes.

The quickest method to address the matching problem is known as "brute force feature matching," and it is described as follows. First, construct a distance function d that compares the descriptors f_i and f_j of two features to determine how far apart they are.

As the two descriptions move closer to one another, the space between them decreases. The distance between each feature in picture two and each feature in image one is then calculated using the distance function d . The feature from image two that is most similar to feature f_i in image one—referred to as our match—will then be returned.

The closest feature to the original one in the descriptor space is this one, which is referred to as the nearest neighbour. The "sum of squared distances or SSD" is the "distance function" that is most frequently used to compare descriptors.

• Sum of Squared Differences (SSD):

$$d(f_i, f_j) = \sum_{k=1}^D (f_{i,k} - f_{j,k})^2$$

Fig.10

DEEP LEARNING BASED ON OPTICAL FLOW:


PERFORMANCE METRICS:

Metrics Name	Purpose	Formula	Expected Value
Point Rotational Error (PRE)	To prevent deviation (error) of a objects angular orientation from an expected, nominal, or commanded angle or displacement.	$PRE = \begin{cases} \cos^{-1} \left(\frac{uu_{GT} + vv_{GT}}{\sqrt{u^2 + v^2} \sqrt{u_{GT}^2 + v_{GT}^2}} \right), & \text{if } (u^2 + v^2) \neq 0 \wedge (u_{GT}^2 + v_{GT}^2) \neq 0 \\ \pi, & \text{if } (u^2 + v^2) \oplus (u_{GT}^2 + v_{GT}^2) = 0 \\ 0, & \text{if } (u^2 + v^2) = 0 \wedge (u_{GT}^2 + v_{GT}^2) = 0 \end{cases}$	Between 0.01 and 0.05
Angular Error (AE)	AE is very sensitive to small displacements, To determine the errors in angles.	$AE = \cos^{-1} \left(\frac{uu_{GT} + vv_{GT} + 1}{\sqrt{u^2 + v^2 + 1} \sqrt{u_{GT}^2 + v_{GT}^2 + 1}} \right).$	Between 1.00 – 2.00
End point Error (EPE)	It gauges the separation between the ends of two optical flow vectors (u_0, v_0), as well as (u_1, v_1)	$\text{sqrt}((u_0 - u_1)^2 + (v_0 - v_1)^2)$	Should have small/min value

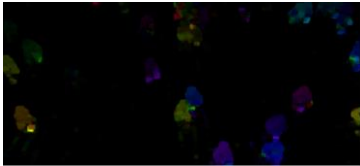
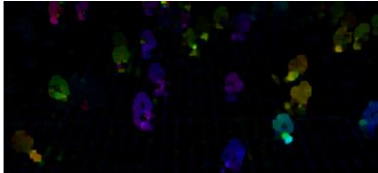
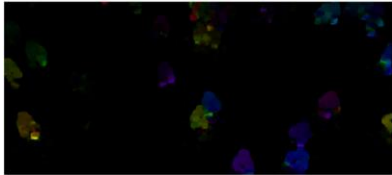
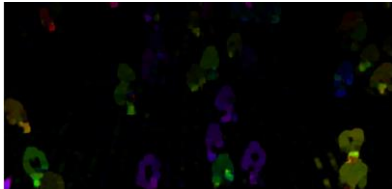
Normalized Euclidean Error (NEE)	The squared distance between two vectors, whose lengths have been normalised to units, is provided. When the vector's direction is significant but its magnitude is not, this is useful.	$NEE = \begin{cases} \frac{\sqrt{(u-u_{GT})^2 + (v-v_{GT})^2}}{\min((u^2+v^2), (u_{GT}^2+v_{GT}^2))}, & \text{if } \min((u^2+v^2), (u_{GT}^2+v_{GT}^2)) > \epsilon \\ \frac{\sqrt{(u-u_{GT})^2 + (v-v_{GT})^2}}{\epsilon}, & \text{if } \min((u^2+v^2), (u_{GT}^2+v_{GT}^2)) = 0 \end{cases}$	Value should be between 0 to 1
Enhanced Normalized Euclidean Error (ENEE)	The relative distance between the E and G T vectors and the use of various normalising techniques are two more ways to overcome the limitations of EPE.	$ENEE = \begin{cases} \frac{\sqrt{(\ P_{GT}\)^2 + \epsilon(\ N_{GT}\)^2}}{\min((u^2+v^2), (u_{GT}^2+v_{GT}^2))}, & \text{if } \min((u^2+v^2), (u_{GT}^2+v_{GT}^2)) > \epsilon \\ \frac{\sqrt{(\ P_{GT}\)^2 + \epsilon(\ N_{GT}\)^2}}{\epsilon}, & \text{if } \min((u^2+v^2), (u_{GT}^2+v_{GT}^2)) = 0 \end{cases}$	Value should be between 0 to 1
Linear Projection Error (LPE)	<p>This metric combines AE and EPE in a way by adding the magnitude of the difference between (u;v) and their perpendicular distance.</p> <p>Based on the perpendicular distance between the two vectors, we may determine the angle between the two non-null vectors.</p>	$LPE = \begin{cases} \ \vec{GT} - \vec{E}\ + \max(\ \text{proj}_{\vec{GT}} \vec{E}\ , \ \text{proj}_{\vec{E}} \vec{GT}\), & \text{if } \ \vec{GT} \cdot \vec{E}\ \neq 0 \\ \ \vec{GT} - \vec{E}\ + \max(\ \vec{GT}\ , \ \vec{E}\), & \text{if } \ \vec{GT} \cdot \vec{E}\ = 0 \end{cases}$	Should have small/min value

a. Lucas Kanade Algorithm:


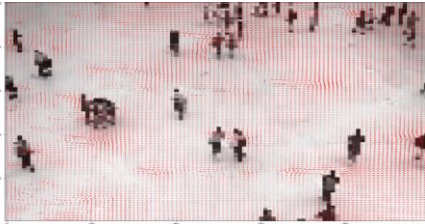


Original video		Time taken for original video is 3mins 41secs
Fast video		Time taken for fast video is 1mins 13secs
Medium video		Time taken for medium video is 2mins 9secs

Slow video		Time taken for slow video is 3mins 23secs
Inference	<p>Lucas Kanade algorithm don't works well in fast video footage, compare to original, medium and slow video footage.</p> <p>Algorithm was able to detect the corners and was able to estimate the where the flow could be in next image sequence.</p>	

b. Dense Optical Flow algorithm:

Original video		Time taken for original video is 3mins 53secs
Fast video		Time taken for fast video is 1mins 21secs
Medium video		Time taken for medium video is 2mins 12secs
Slow video		Time taken for slow video is 3mins 47secs
Inference	<p>Dense Optical Flow algorithm works well in fast video footage and Slow video footage, compare to original and medium video footage.</p> <p>Dense optical flow able to compute optical flow vector for each and every pixel of each frame.</p>	

c. Horn and Schunck:

Original video (Frame 137)	
Fast video (Frame 35)	
Medium video (Frame 71)	
Slow video (Frame 50)	
Inference	As the people in the video don't have constant velocity movement the estimation in all the footage is little poor, which result in error at the boundaries.

In Horn and Schunck, as we know it can only handle object/person/things which has constant velocity. In our dataset we don't have any constant movement kind of objects/person/things, which resulted in poor estimation.


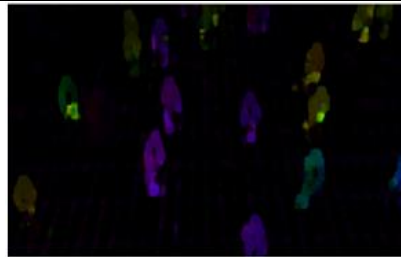
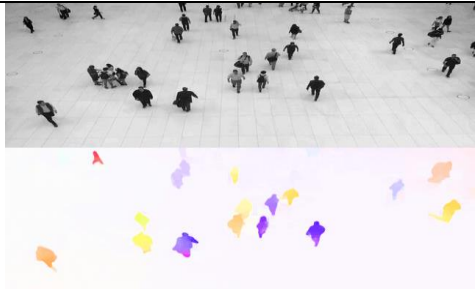
In Lucas-Kanade, this algorithm works fine with slow motion footages/recordings. And it don't work for fast motion footages/recordings. We can clearly see in our output, where we got better estimation of flow in next image sequence in slow motion footage then fast motion footage. In fast motion footage, the estimation was mismatched and uneven.

In Dense Optical Flow, as it compute optical flow vector for each and every pixel of each frame. We where able to estimate the flow in all the footage clearly. Even though the computational time is much higher then other two algorithm. Its give accurate result and denser result which is suitable for application like motion and video segmentation.

From comparison, Dense optical flow was able to give better estimation to describe image motion.

RAFT:

Estimate the motion of the image intensities over time in a video using optical-flow algorithms like RAFT. i.e. to calculate Pixel points within the images, and to provide the estimation of where the points could be in the next image sequence.

Optical Flow Algorithm	
Dense Optical Flow Algorithm	
RAFT Algorithm	

FACE DETECTION:

Face recognition using A machine learning technique called Haar cascades involves teaching a cascade function from a set of input data. We'll make use of OpenCV's face classifier, which comes with a lot of pre-trained classifiers for faces, eyes, smiles, and other features.

a. Methodology:

- Taking in the supplied picture data.
- Creating grayscale versions of the input pictures.
- Applying the Haar cascade is step iii.
- Assessing classifiers in terms of accuracy and processing speed.

A. Bringing in the necessary libraries

B. Using the pictures that the camera has taken.

C. The image is transformed into a grayscale image before being processed by the classifiers.

D. OpenCV will be used to load the image.

E. The BGR colour space will be used by default when loading a picture.

Cascade classifier in Haar

- Using the built-in method `cv2.imread(img path)`, which here accepts the image path as an input argument, to load the input image
- Gray-scale mode conversion, followed by display

iii. the Haar cascade classifier being loaded

•Pixelvalue=(SumoftheDarkpixels/NumberofDarkpixels)(SumoftheLightpixels/NumberofLightpixels) is a formula for calculating the Haar value.

The Haar cascade achieves accuracy of 96.24 percent.



Fig.11. Input



Fig.12. Output

V. RESULT :

The suggested system is designed to detect objects, people, and activities. The scene's video is recorded by the camera and then translated into frames by the CPU. It is important for the user to be aware of everything around them. The voice output is generated, and the system uses audio output to direct the user to the item. image detection and user guidance for what is in their path.



Fig.13. Input

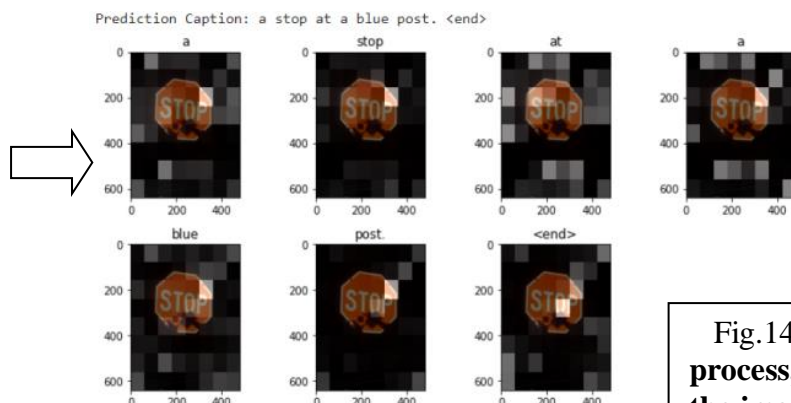


Fig.14. After system process, we get captioning of the image, describing what is there in it.

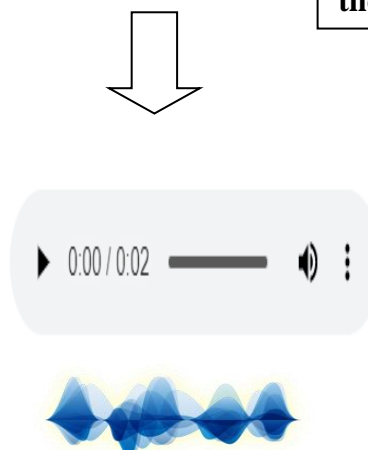


Fig.15. Audio Output Generation

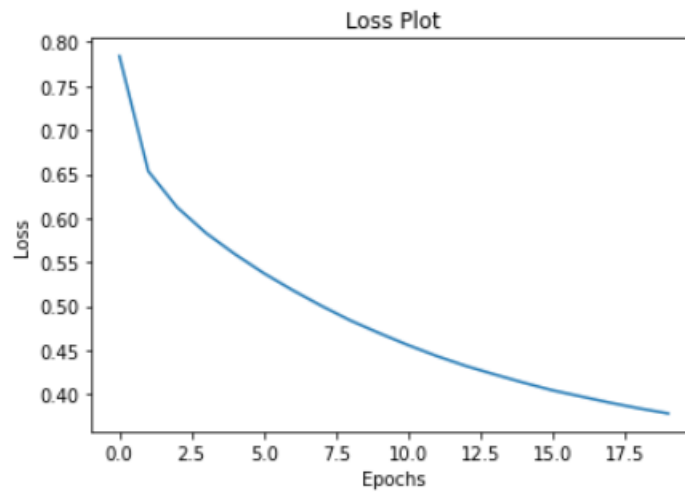


Fig.16. Epochs VS Loss

VI. CONCLUSION:

This research describes a cutting-edge method for helping persons who are blind. The suggested system's straightforward architecture and user-friendliness enable the subject to be independent in his or her own house. The technology also seeks to assist the blind in navigating their environment by seeing obstacles, finding their essentials, and reading signs and messages. Initial tests have produced encouraging results, allowing the user to securely and freely move about his environment. Voice Assisting makes the system considerably more user-friendly.

VII. REFERENCES:

- [1] B. Deepthi Jain, S. M. Thakur and K. V. Suresh, "Visual Assistance for Blind Using Image Processing," 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 0499-0503, DOI: 10.1109/ICCSP.2018.8524251.
- [2] P. Vyavahare and S. Habeeb, "Assistant for Visually Impaired using Computer Vision," 2018 1st International Conference on Advanced Research in Engineering Sciences (ARES), 2018, pp. 1-7, DOI: 10.1109/AREX.2018.8723271.
- [3] M. P. Arakeri, N. S. Keerthana, M. Madhura, A. Sankar and T. Munnavar, "Assistive Technology for the Visually Impaired Using Computer Vision," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018, pp. 1725-1730, DOI: 10.1109/ICACCI.2018.8554625.
- [4] Y. Zheng, M. Zhang and F. Lu, "Optical Flow in the Dark," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 6748-6756, DOI: 10.1109/CVPR42600.2020.00678.
- [5] Y. Zhang and F. Wang, "Improved Optical Flow Algorithm of Moving Object Detection," 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 2015, pp. 196-199, DOI: 10.1109/IMCCC.2015.48.
- [6] D. K. Liyanage and M. U. S. Perera, "Optical flow based obstacle avoidance for the visually impaired," 2012 IEEE Business, Engineering & Industrial Applications Colloquium (BEIAC), 2012, pp. 284-289, DOI: 10.1109/BEIAC.2012.6226068.
- [7] W. Tarimo, M. M. Sabra and S. Hendre, "Real-Time Deep Learning-Based Object Detection Framework," 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 1829-1836, DOI: 10.1109/SSCI47803.2020.9308493.
- [8] G. Chandan, A. Jain, H. Jain and Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1305-1308, DOI: 10.1109/ICIRCA.2018.8597266.