Walking Assistant / Traffic Sign Detector for the Visually Impaired

Hlulani Trevor Nkuna
Computer Systems Engineering
Department Faculty of ICT
Tshwane University Of Technology
Pretoria, South Africa
trevornkuna7@gmail.com March 2021

Abstract - Having witnessed the impact of the blindness disability on people's daily lives and routine, this document discuses a technique suggested in order to improve life experience of visually impaired people. The motivation of this research is that the market is not currently offering such products at an affordable price, meaning only a selected few can afford to own it. This project focuses on machine learning algorithms that help improve the user's experience by recognizing road signs, objects and obstacles in the path of the user. Lenet convolutional neural network model is used for training for classification of different traffic signs. We train the model using a German Traffic Sign Recognition dataset from Kaggle. The number of epochs used is twenty five iterations, and the experimental results produce an accuracy of ninety seven percent. The processing time per frame is relatively low i.e signs are detected quicker and it's accuracy is much higher, this indicates that the product is safe and more convenient for everyday use.

Keywords-*traffic sign recognition, intelligent vehicles, traffic sign detection, convolutional neural network, epochs, Ultrasonic, Exploratory data analysis, pre-processing, architecture.*

I. Introduction

The history of artificial intelligence dates back to antiquity with philosophers mulling over the idea that artificial beings, mechanical men, and other automatons had existed or could exist in some fashion. Artificial intelligence is a subset of computer science that focuses on machine-driven intelligence (i.e. non-human intelligence). AI is the understanding that machines can interpret, mine, and learn from external data in a way where said machines functionally imitate cognitive practices normally attributed to humans. Artificial intelligence is based on the notion that human thought processes have the ability to both be replicated and mechanized.[9]

A recent study by the World Health Organisation has revealed that approximately 285 million people are affected by visual impairment,214 million have a slight vision and 39 million of which are blind.[6] This is a major problem as it limits one's movements and freedom, i.e. you making it difficult to move from point A to point B on your own. This use of walking canes has been made available but it is only helpful for detecting physical objects in front of the user. Smart canes have also been made available with a wider range of functionality but this comes at an expense and thus not easy for the public to acquire. In a recent study done by Mobility Device Statistics in the USA on 6.8 million people who depend on mobility devices, 1.7 million people use wheelchair, scooter riders depend on crutches and canes. This means that a vast majority of individuals use traditional canes and their day to day activities are limited due to having to need assistance to travel or walk to unfamiliar territories. Other types of walking assistants include dog guides, 3sonar canes.[6] Dog guides can also become expensive to acquire and also the fact that they are animals, they may still need to be taken care of, bathed and taken to the veterinary doctor which can prove costly over time. Sonar canes are limited due to the fact that they may only sense between eye and foot level and only to a limited distance, this becomes an issue if you're in a busy road where there are traffic lights and different traffic signs. Since it can not detect these traffic signs, this may lead or result in accidents and fatalities.

The Walking assistant we are focusing on in this paper combines two solutions. The first part is the object detection, using sonar sensors, this is to assist when walking in public or in tight spaces. It uses it's echo and trig pins to detect obstacles, it could be people or objects. It then sends feedback to the user by a certain beep, the user must then navigate through an open path i.e. without obstacles. The second part is to detect different traffic signs so this can assist the blind user when approaching intersections or crossings. The system is trained on 43 different traffic signs using the German Traffic Sign data-set. When a sign is detected, a voice feedback using Google Text To Speech library is applied to notice the user.

These two combined solutions provide the user with an advantage over the other solutions in the sense that it gives the user a more visual feel, it also eliminates the need of human assistance to complete general tasks.

II. Literature Review

In recent times, Recognition, Classification and Detection of Traffic Signs has become a significant not only for Walking assistance to the blind but also for autonomous driving. The evolution of Computer vision has made this easier to study and implement through different complicated models. Convolutional Neural networks is at the core of this study and allows for different methods to be explored, implemented and improved upon. In [5], the authors propose an Intelligent Transport system that makes use of sensors to identify road signs using two main steps i.e. detection and recognition. In the detection phase, the image is pre-processed, enhanced, and segmented according to sign properties e.g shape and color. The results is an image that is segmented into regions that can be recognized as possible signs. The difficulties that can be faced is that as the image is segmented into regions, the colors can be affected due to surroundings and background, which will therefore result in inaccurate outputs. In [7], the authors propose a Traffic Sign Detection and Recognition System also based on Color Segementation, Shape matching and in addition an SVM. Several methods are involved here, Image Acquisition (Images collected from a board camera) .Image preprocessing is then peformed to remove the low[1]frequency background noise, normalising the intensity of individual particles images, removing reflections and mask the portion of images. Shape matching based detection is peformed to accelerate the proceedure without employing model based classifiers. Object feature analysing is used to eliminate noise from 4the image to better deal with ROI.Shape Matching and candidate selection. As almost all signs containing red color are round and octagonal, the proposed method drew on these common shapes to detect hypothetical shapes which are close to traffic signs. After the detection of traffic sign, the region of interest (ROI) is passed to the SVM for recognition[7]. Some of the drawbacks to this problem is the partial obscuring, multiple traffic signs appearing in a single frame, the blurring and fading of traffic signs. In[3], a detection and identification of traffic signs for driving assistance systems is proposed. They first filter out most of the non sign parts of an image using the color information. Regions that may have image blocks are then extracted and further extract regions from the above image blocks. A convolutional nueral network is applied to verify the candidate areas of non-road signs and then identify the traffic sign. In [1], they propose a system that can be fitted into cars. It makes use of deep learning algorithms(CNN), feature extraction and classification. The drawback here is the speed in which it pefeorms i.e reaction time for the image to be classified. Another drawback is that if multiple signs are detected then it cannot properly classify the image.In [2], a traffic sign detection and recognition algorithm is suggested. It uses the Lenet-5 CNN model to effectively classify images based on their traffic sign classes, this system works to identify these in real time. The model has the following short commings. Some of these are that the original convolutional kernel does not perform well in feature extraction, as a result the extracted features cannot be properly used for the accurate classification.

The above mentioned algorithms work extremely well when it comes to clas[1]sification, all with an average accuracy of 80% or more. However, these can be improved upon as some of the comming issues faced are feature extraction, background noise elimination and speed/accuracy in real-time. To reduce these errors we will take a look at the Lenet-5 model architecture, which performs well in real time, it is a little bit slower when compared to the moden CNN models but it this means that it detects objects more accurately[10]. We will also make use of the pretrained models which are likely to be more accurate due to the increased number of epochs when training. In addition to the traffic sign we will add more common real-life objects to our dataset this includes, cars,trucks, people, etc. So that our system detection is not limited to traffic signs only. This can help reduce the number of blind-pedestrian accidents[4].

III. Methodology

Taking note that we implement and combine two solutions into our system, we will discuss the methodology to implement and test each. In the first method we use an arduino Uno, 2 x ultrasonic sensors(SR-H04), jumper wires, 1x buzzer, 2x LED's. We use the mentioned apparatus to build an ultrasonic system that makes use of two sensors to detect obstacles ahead of the user. The ultrasonic sensor has 4 pins, VDD, GND, Trig, Echo. VDD and GND are the power and ground of the circuit relatively. Trig is the output pin that sends an ultrasonic signal, Echo is the input pin that takes in the feedback of the signal sent out by 5the Trig pin. We then construct a formula to determine the distance, we make use of the time taken for the echo to receive input after the trig has initial sent out a signal.

The formula is as follows:

distance = (time * 0.034) / 2.

This will give us the distance in centimeters and based on this information we can take a decision to alert the user if there's an obstacle ahead at a distance of 40 cm or less, This implies that if the user continues walking in the same direction they will collide with the obstacle. The picture of the circuit is shown below in figure 2 and figure 3 indicates the flowchart of the whole process.

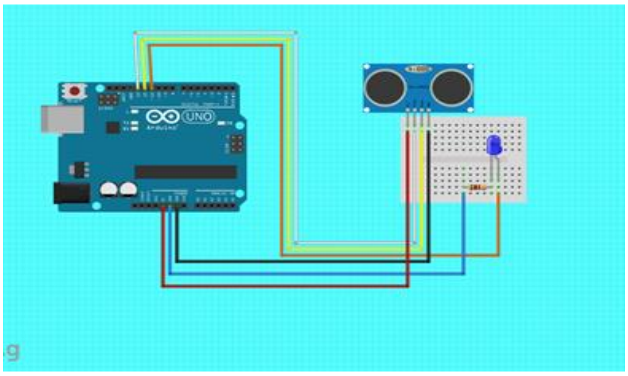

Figure 1: Circuit diagram of Ultrasonic Sensor

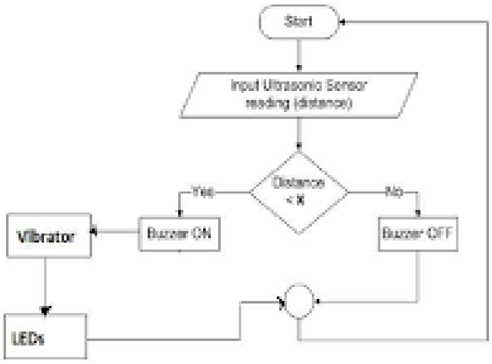

Figure 2: Flow chart of Ultra Sonic Sensor system

Figure 3:Stop Sign(Sample image).

Figure 4 : Sample image(Traffic Sign).

The second solution is to make use of Artificial Intelligence Convolutional Neural networks for computer vision. We use the Lenet-5 architechture model for image classification. We first collect training data i.e. Traffic Sign images data-set. In this case we use the German Traffic Sign Dataset from Kaggle. We read these images into our note book and perform EDA(exploratory Data Analysis), we convert the images to 32x32x1(as required by the Lenet-5 6model), we then perform data augmentation i.e. rotate the images to different angles, different scale colors(e.g grayscale).[8] The Lenet-5 receives and input image of 32x32x1 so it can better recognise digital patterns. It uses 5x5 filter and a stride of 1.[8] The model is trained using approximately 25 epochs(iterations).

• W x H → 32 x 32 (Width x Height)

• F(w x h) → 5 x 5 (Filter)

• S → 1 (Stride) P → 0 (Pooling)

= W − F w + 2P Sw2 (1)

The next layer is the pooling layer, and the formula to calculate the pooling layer is as follows.[8]

• IM → 28 (Input Matrix → Convolution output volume., Refer above derivation output)

• P → 0 (Pooling)

• S → 1 (Stride) S → 1 (Stride)

= IM + 2P − 2 S (2)

The last stage is the Fully Connected Layer. It has 120 nodes and then followed by another Fully connected layer that has 84 nodes. It uses Sigmoid and Tanh non-linierity functions[8].Sigmoid and Tanh functions are activation functions and are shown below :

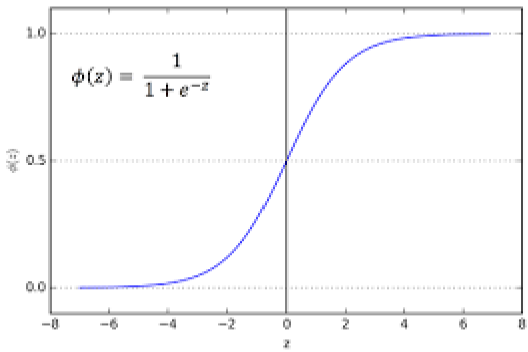

Figure 5: Diagram of Sigmoid activation function

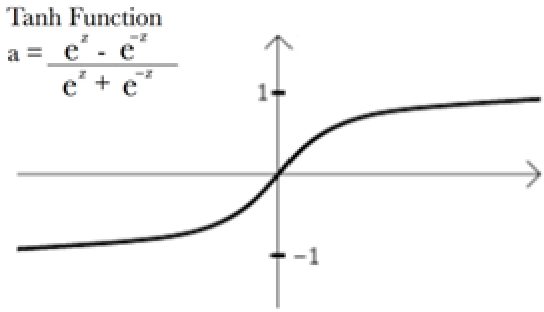

Figure 4: Diagram of Tanh activation function

The diagram of the Lenet-5 is shown below[8]:

• Designed in 1990's to identify handwritten digits in MNIST data-set

• input images = 32*32*1 gray scale

• followed by two pairs of Convolution layer with stride 2 and Average pooling layer with stride 1

• Finally, fully connected layers with Softmax activation in output layer

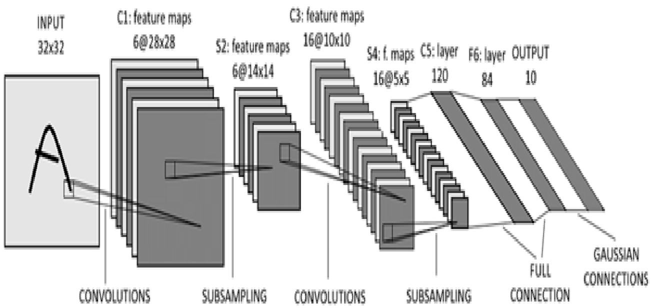

Figure 6: Diagram of Lenet-5 architechture

IV. Results and Discussion

The ultrasonic object detection was implemented and tested physically indoors and also out doors, it has resulted in a 100% accuracy meaning that for as long as it is positioned properly, it detects objects 100 percent of the time. The test was done 9 times, the formula and the results are calculated below:

Accuracy = (TruePositive + True negative) / (True Positive + True negative + FalsePositive + FalseNegative)

The Convolutional Network System achieved an overall test accuracy of 97.5% in the classification and 96.5% in the detection part of the experiment. After reaching 30 epochs, the accuracy started producing a constant value of 99.32 and we should take note that the test score was 99.32% and the loss was 2.39 % as indicated in both figure 7 and figure 9. We also notice that for each class, the more images we have the more accurately the system predicts, this relationship is directly proportional as indicated by figure 8.

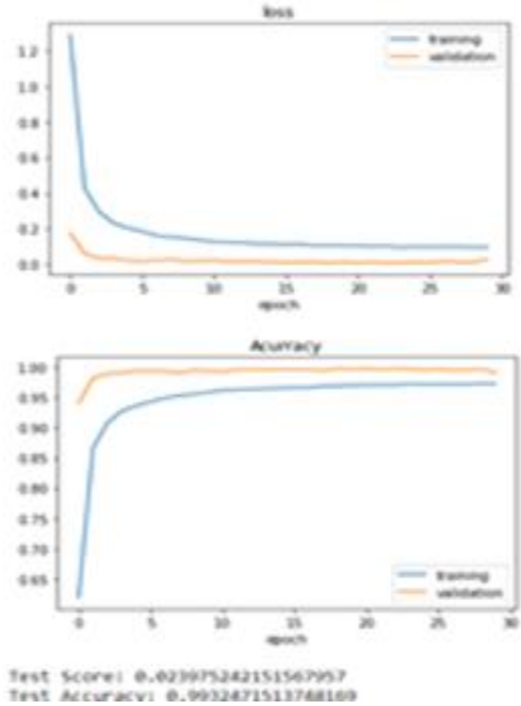

Figure 7: Accuracy and loss of the training data

The Confusion Matrix also helped us visualize how model discriminates between classes. Allowing us to calculate the accuracy, precision and recall as showing in the equation and ground table below[8].

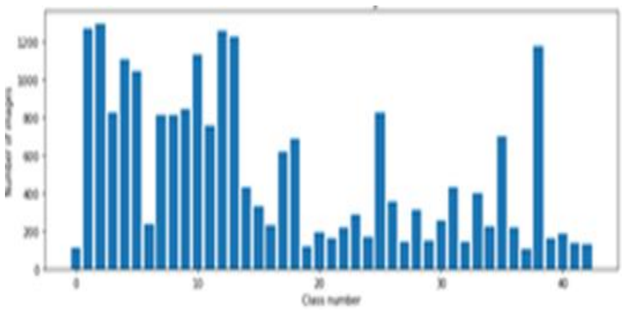

Figure 8: Histogram of all the image accuracy

Number of Images	Number of Iterations	Accuracy
20000	5	45.68
20000	10	60.37
20000	15	65.32
20000	20	85.29
20000	25	97.56
20000	30	97.62

Figure 9: Histogram showing the accuracy per iteration.

• True Positive: Number of Positive entries correctly classified

• False Negative: Positive samples predicted as Negative

• False Positive: Negative samples predicted as Positive

• True Negative: Negative samples predicted as Negative

Accuracy = (TrueP ositive + Truenegative)/(TrueP ositive + Truenegative + FalseP ositive + FalseNegative)

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Figure 10: confusion matrix table

V. Conclusion

In this paper we have looked at an alternative way to assist visually impaired people through the implementation of a smart walking guide that helps them detect objects using a simple circuitry with ultrasonic sensors, it also makes use of Neural Networks(convolutional) to aid in the detection of different traffic signs(total of 43 classes). The aim was to better the experience of visually impaired people and give them the flexibility to be able to perform more and more tasks by themselves without needing human assistant. The test results indicate that this is a step in the right direction as the methods implemented in the paper have a relatively high accuracy and have also proven to be much more cost effective(cheaper). This model can be further optimized and improved upon in the future as it has shown that it lacks in some aspects such as speed when detecting objects. Faster models such as YOLO v5 have since been implemented which may be more faster but the disadvantage would be that you trade accuracy for speed and since the loss in accuracy is numerically insignificant, this can be considered as an improvement.

References

[1] Danyah A. Alghmgham. Autonomous Traffic Sign (ATSR) Detection and Recognition using Deep CNN. Science Direct, 2019.

[2] Jingwei Cao. Improved Traffic Sign Detection and Recognition Algorithm for Intelligent Vehicles. MDPI, 2019.

[3] Chin-Chen Chang. Traffic Sign Detection and Recognition for Driving As[1]sistance System. Hueng-Yin Lin, 2018.

[4] Alper Fatih Yilmaz. Object Detection and Tracking. 2012..

[5] Hasan Fleyeh. Road and traffic sign detection and recognition. Research Gate, 2006.

[6] Md. Milon Islam. Developing Walking Assistants for Visually Impaired People: A Review. Research Gate, 2019.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[7] O Safat B. Wali. An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation, Shape Matching, and SVM. Safat B. Wali et al, 2015M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Prabhu. CNN Architectures — LeNet, AlexNet, VGG, GoogLeNet and ResNet. towardsdatascience, 2018.

[9] Redd Rebecca Reynoso. A Complete History of Artificial Intelligence. Rebecca Reynoso, 2019.becca Reynoso. A Complete History of Artificial Intelligence. Rebecca Reynoso, 2019.

[10] Javier Rodrigo Ruiz-del Solar. Object Detection: Current and Future Directions. Frontiers in Robotics and AI. 2015.