**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

# Automated Voice Assistance for Visually Impaired People Using Deep Learning

**Dr. J. Preetha[1], M. Ganthimathi[2] K. Kamaleshwaran[3],**
[1]Professor,[2]Associate Professor,[3] PG Scholar,
[1,2,3]Department of Computer Science and Engineering Muthayammal Engineering College, Namakkal, Tamil Nadu, India
psgpreetha@gmail.com[1], gandhimanimay02@gmail.com[2], kamalkandhs@gmail.com[3]

**Abstract**

Vision loss in the elderly is a major health problem. In past years, the main choice of visually impaired people is a spectacular glasses. Today's assistive technology tools include not only better magnifying devices, but also advanced computing software, applications, and other use cases that use facial recognition, artificial intelligence, and other technologies to visibly improve the lives of the visually impaired. In these modern days, many developments have been madeto improve the quality of life of the visually impaired. These people face many challenges in performing their daily tasks, including discovering themselves in an unfamiliar environmental circumstance. They cannot distinguish the objects around them and cannot accurately perceive the objects that around them most of the time. Keeping this in mind, With the advancement in the deep learning frameworks like computer vision techniques assist the visually impaired. In this paper, our goal is to providing the automated way that the people can get the knowledge of the facing objects in real-time by voice feedback. And also they get the position of the objects in front of them. We adopted the Mobilenet Single-Shot Detector, a deep learning multi-label detection model for classifying theobjects. Besides, the model determines the position of object in the frame and finally generates an audio signal as output to notify the visually-impaired people. This system provides a novel way of assisting visually challenged people.

**Keywords— object classification; Automate the vision process; Single-Shot Detector Model;pyttsx3; GTTS**

## INTRODUCTION:

Visually impaired people face many difficulties in their daily routines. Current statisticspublished by the World Health Organization(WHO) 2019 shows that all over the world, there are more than 2.2 billion visually impaired people around the world who awaken in eternal darkness. For most people, their eyesight is vague or blurred.Of these, 62 million live in India, which is partially or completely blind. A lot of research has been done to deal with the inconveniences of daily life, and as a result, various amenities for everyone have been provide. However, there are still many inconvenient for the visually impaired. The greatest inconvenience experienced by a blind person in Daily life consists of finding information about objects in indoor and outdoor. But now they are supported by technology that makes everyday life easier.

One of the biggest inconveniences that blind people experience in their daily lives is finding information about outdoor environment. Previous studies included object analysis using ultrasonic sensors and other edge technologies. However, with these methods, it is difficult to know exactly where the object is, especially in the presence of obstacles. In this article, we will get accurate object information and use deep learning object detection techniques to obtain location. Advances in technology, especially the rise of computer processing features such as deep learning (DL) models and the emergence of wearables, are paving the way for the visually impaired. Previously specifically designed for the visually impaired, models work well to detect a single object. However, in real-time scenarios, these systems do not provide effective results for the visually impaired. In addition to object detection, additional information about the location of objects in the scene is essential for the visually impaired. Against this background, the current research introduces an efficient object recognition model equipped with a voice support system by audio generation using pyttsx and GTTS a python cross-platform libraries.

The characteristics of an object such as edge, shape, and intensity are necessary to recognize objects in an image using object

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

detection technology. SSD mobilenet framework model with TensorFlow is used to detect objects. In addition, it is equipped with a speech synthesizer, so that the recognized object can respond to the visually impaired peoples as a voice output.

**LITERATURE REVIEW:**

**Kartik Umesh et al., "A review and approach for object detection in images" in the year January 2017**

The basic input of an object detection system can be an image or a scene in the case of a video. The raw objective of this system is to detect the objects present in the image or scene or in other words, the system has to distinguish different objects into respective feature classes. The object detection system consists of two main phases, namely: the learning phase and the testing phase. Learning by training mainly consists of learning block where the appropriate learning model is defined, it can be piece based or patch based etc. Then, the object model block uses the operations that were done previously to represent the objects with various representations such as histogram representation, random forest representation, etc. On the other hand, learning validation blocks doesnot require any kind of training because they are pre-validated. Therefore, after preprocessing theimage, an exact class match is performed to generate the features of an object in the image. Themain objective of the testing phase is to decide if an object is present in the image provided to the system as input and, if so, to which class offeatures it belongs. The image is then queried for an object using various search techniques such as sliding window technique and according to thesearch engine output, a decision is made on the feature class. If training because they are pre- validated.

**Wei Liu et al., "Single-shot Multi-Box detector"in the year December 2016**

The SSD approach is met by a feed agglomeration network that generates a fixed-size set of bounding boxes and scores for the presence of feature class instances in those boxes, then followed by a zero-maximum deletion step to provide the final detected result. The first networklayers are based on a standard architecture used forhigh-quality image classification (truncated beforeany classifiers), which we will call the underlying

network. Multi-scale feature maps for sensing additional layers of complex features at the end of the truncated baseline network. These layers are decremented and allow detection prediction at different scales. A version of it is also required fortraining in YOLO and the regional recommendation phase of Faster RCNN andMultiBox. Once this assignment is specified, the loss function and reverse propagation are applied end-to-end. Training also includes showing the default set of boxes and scales for detection as wellas strategies for extracting and supplementing negative data.

**Daniel Fleury et al., "Implementation of Regional-CNN and SSD Machine Learning Object Detection" in the year November 2016**

TensorFlow Object Detection API, published by Google, is an open source framework for object detection related tasks used to train Single Shot Detector (SSD) and regional- Convolutional Neural Network (R-CNN) models since their goal is to provide scalability and potential for Google device deployment. Most importantly, Google has prepared TensorFlow engines with necessary support for mainstream methods like MultiBox/SSD and Fast/Faster RCNN. Object Detection APIs are created with various levels ranging from implementations tosimple box operations. The lowest API level typically includes box operations, box representation, target assignment, and visualization techniques. The higher API level covers the core structure of the super architecture including SSDs,faster RCNNs, and more. The benefit of implementing a faster TensorFlow RCNN model is that the training time required to generate a checkpoint file with acceptable loss values is halved compared to using the portable SSD model.Net TensorFlow. In addition, it detects a higher number of objects per frame as well as increases theaccuracy for detection in comparison. However, this idea cannot be used in real-time object detection analysis since there is no convolutional layer and deep point layer, which also has a lower detection rate.

**Sheng Ding, Kun Zhao "Object detection basedon a deep neural network" in the year November 2018**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

The development of the detection algorithmis divided into two phases. The first stage is based on the traditional functionality of the solution and the second stage is a deep learning algorithm. Witha large amount of detection data, the techniques oftraditional detection methods will become saturated. The detection performance will increasestep by step, but the improvement will decrease after a certain amount of data. However, with deeplearning methods, it is different. As scene deliverydata accumulates, detection performance continuously improves.

A set of daily provisioning data is collectedand then various training object detection models are applied to the data. And by comparing direct training and parameter tuning model training, it will be demonstrated that the convergence speed and accuracy of object detection are improved by tuning the parameters.

**ALGORITHMS AND TECHNOLOGIES:SSD MOBILE-NET MODEL:**

On the account of selecting this method for detecting objects in images using a single deep neural network. Because it customizes the bounding box output space to a simple set of default boxes on different aspect ratios and scales for each feature map location. In real-time prediction, the network generates a score for the presence of each category of object in each defaultbox and makes adjustments to the box to better fit the shape of the object based on the threshold value. In addition, the system meets the expectations of many different-purpose entity maps for normally handling objects of different sizes. Our SSD modelis a direct comparison to object recommendation request strategies, as it delivers with recommendation generation and the next pixel or feature resampling step and encapsulates all the features. computation in a single model. This makes SSDs simple to prepare and orchestrate in frames that require ID segments. The test results onthe PASCAL VOC, MS COCO and ILSVRC
datasets confirm that the SSD has almost identical accuracy to techniques using the extra article recommendation step and is much faster while stillproviding the same configuration. architecture for both training and inference. Unlike other single- stage techniques, SSDs have much better accuracyeven with small information image sizes. For
300×300 frame data, the SSD delivers 72.1% mAPon the 2007 VOC test at 58 FPS on the Nvidia Titan X, and for 500×500 data, the SSD delivers 75.1% mAP, a rating higher level failure than the faster RCNN model.

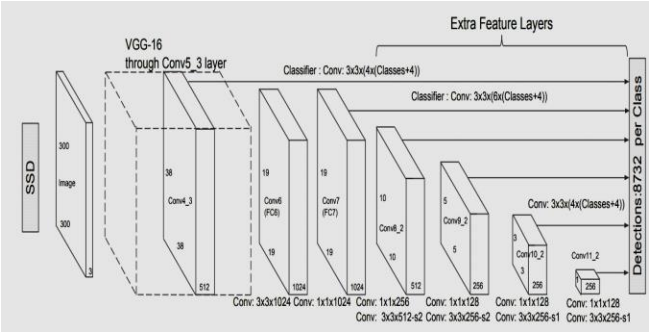


**Fig. 1: SSD model architecture**

For mobile and embedded image processing applications, we used a class of effective models called MobileNets. MobileNets isbased on a streamlined architecture that uses deeplyseparable aggregates to build lightweight deep neural networks. We introduce two simple global hyperparameters that effectively balance latency and accuracy. These hyperparameters allow model builders to choose the right sized model for their application on basis of the constraints and use casesof the problem. We present extensive tests of the balance between resources and accuracy and show strong performance compared to other popular models on the ImageNet classifier. It further demonstrates the viability of MobileNets across multiple uses and use cases, including object detection and classification, fine grain classification, face attributes, and large-scale geolocation.

**GTTS AND pyttsx3:**

Google has developed Google Text to Speech, a screen reader application for the Androidworking framework. Enable a read-through application to allow anyone to hear (speak) the content on your screen using some dialects.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

Content-to-speech can be used in applications such as: For example, Google Play Books for reading books, Google Translation for reading echoing interpretations that provide valuable information on how to express words, Google Talkback and other availability based voice-to-speech applications, and third-party applications. Converts content into human-like audio and uses over 180 audio in over 30 dialects and variations. Applying

core research on speech synthesis (WaveNet) and Google's incredible neural system to provide high-fidelity sound. Integrating Selective Access to WaveNet Innovation DeepMind conducts basic research on AI models to create discourses that copy human voices and sounds with increased regularity and reduce holes in human execution by 70%. bottom. Cloud Text to Speech provides elite access to over 90 WaveNet voices and will be phased out over time.

During the audio signal generation step, the detected object and its position in the frame converted to an audio signal using the Python library pyttsx3. Be cross-platform texttospeech conversion library, it's platform independent. Furthermore, a great advantage of this library is that it also works offline. The python code snippet of using pyttsx3 is provided below.

**METHODOLOGY:**

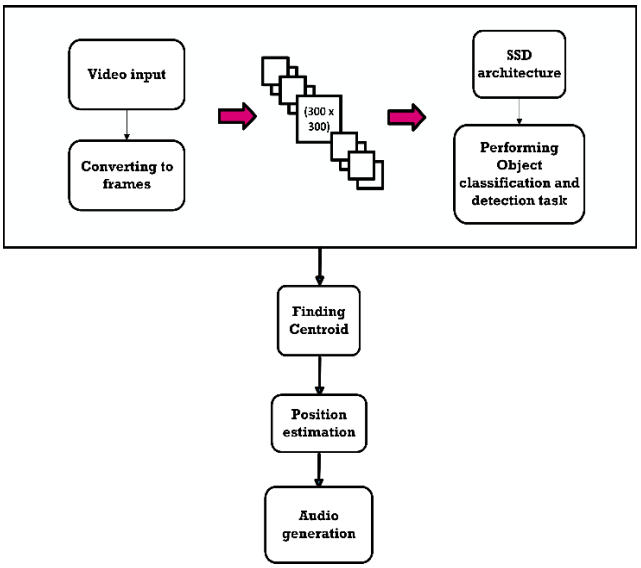


**Fig. 2: System workflow**

**Setting up the project environment:**

The deep learning environment of the working system have to include Anaconda, an opensource scalable software program solely for python, Machine learning, and other data science frameworks know-how with appropriate CUDA, cuDNN packages, a GPU overall performance accelerator. With those packages, we use TensorFlow-GPU, a giant deep learning library to system a huge quantity of tensors and Microsoft visual studio, an integrated development environment to act and perform with deep learningoperations.

**Dataset preparation:**

A large number of images of an object is required by Tensorflow to implement the best detection classifier. To train the classifier, the training images must have random objects in the dataset as well as suitable objects and must have a variety of backgrounds and lighting conditions. There should be multiple images where the relevant subject is partially visible, overlaps with somethingelse, or is just in the center of the image.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

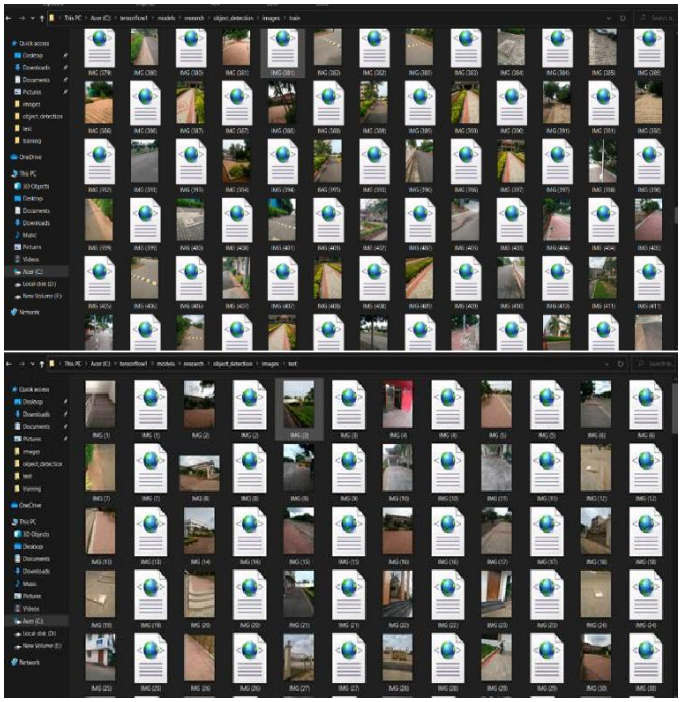

**Fig. 3: Training and testing images along with .xmlfiles**

For my object detection classifier I have 8 different objects like road, path, damaged path, tree, light pole, door, building, steps, because we take These objects make a data set according to thedaily obstacles met by the visually impaired. We used our phone to take about 1100 photos and manually labeled them using the labelimg tool.

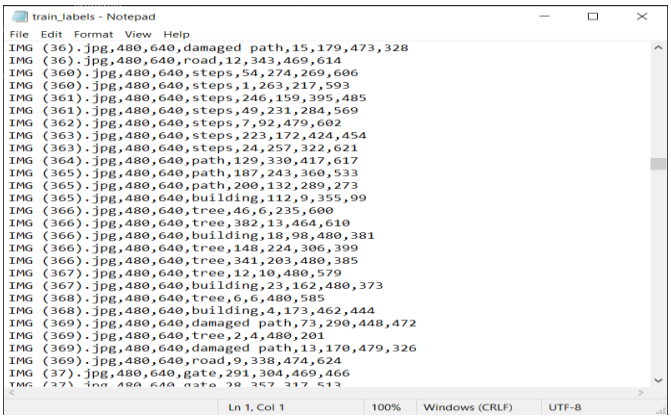


**Fig. 4: Dataset in form of .csv file format**

LabelImg is a free and open source tool forgraphic labeling of images. It is written in Python and uses QT for its GUI. It's an easy and free way to tag a few hundred images to try out your next object detection project. Labels are used to help identify the components of your data that you wantto train your model to identify in the unlabeled dataset. High-quality datasets are essential for computer vision and high-performance modeling. Computer vision modeling follows the garbage,garbage philosophy, which means it's important tolabel images carefully and correctly. We've created a labeling guide to help ensure your training dataset is of high quality. Correctly labeled data is critical to the success of machine learning, and computer vision is no exception.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

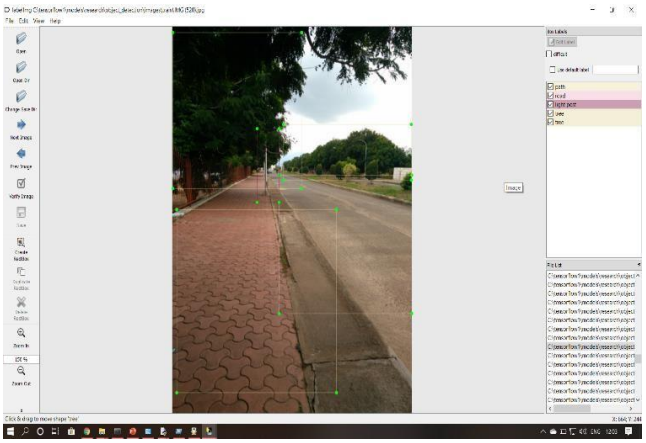

**Fig. 5: LabelImg tool**

**Training Stage:**

With the labeled images, it's time to generate TFRecords words that function as input tothe Tensorflow training model. At this stage, the classifier can learn the correspondence between theinput and output parameters. Next, the object detection training pipeline must be configured. It includes an SSD model and several settings to saveeach model's key checkpoints for later use. Every stage of formation is met with loss. It will start at ahigher value and decrease as the workout progresses. For us training on the mobilenet-v3 portable SSD model, it starts at around 36.85 and will train consistently under 3.444.

With the help of Tensorboard we can see the overall training progress. An essential graph is the loss graph, which shows the overall loss of thetraining model over time.

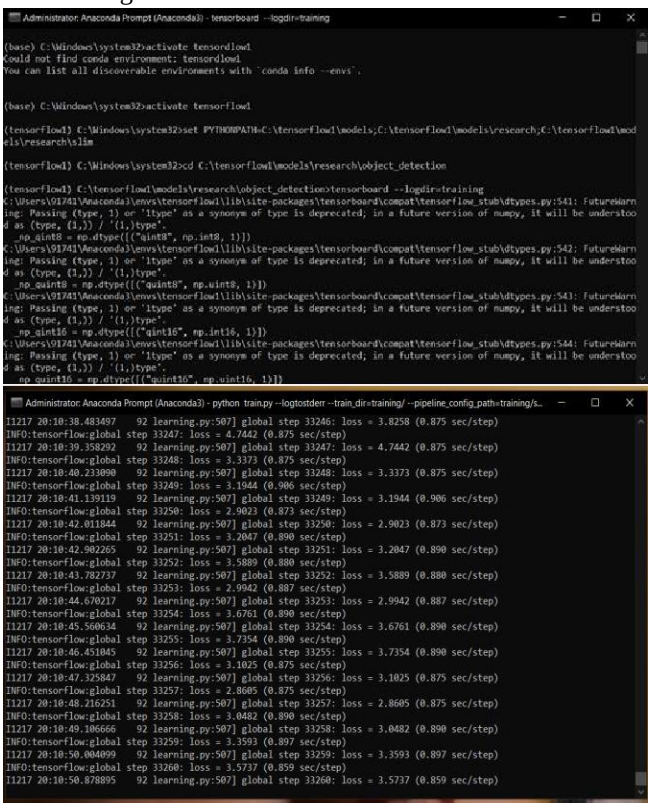


**Fig. 6: Training pipeline**

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

**Generating an inference graph:**

After the training stage completion, the next step is to generate the frozen inference graph (. pb file). This protobuf file consists of the object detection classifier. The frozen graph is the method to perceive and store all the graphs, weights, etc... in a one unique file that we are able to effortlessly use.

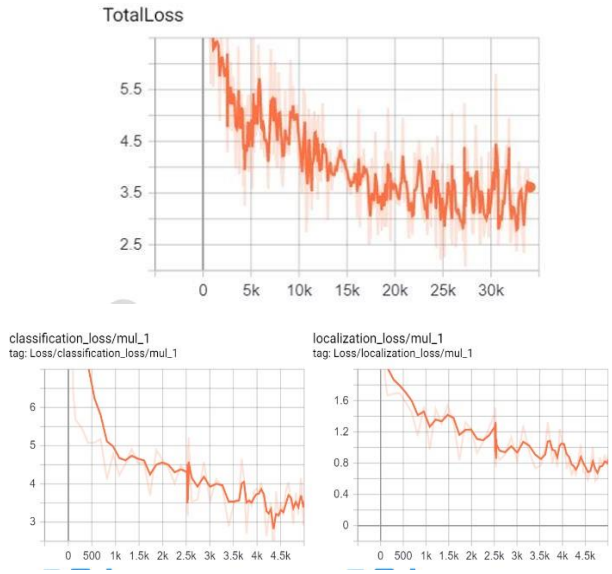


**Fig. 7: Monitoring Total loss, Classification andLocalization loss graphs**

**Finding position of an object:**

Once the SSD model identifies the objects in the frame, the next step is to define the position of objects. For this, each frame is decomposed into a 3-row x 3-column grid cell as shown in the image below. The whole image is divided into three positions like top, middle and bottom as a row and left, center and right in the column. Then the center position of each bounding box is calculated based on the coordinates of the box like x, y, width (w) and height (h).

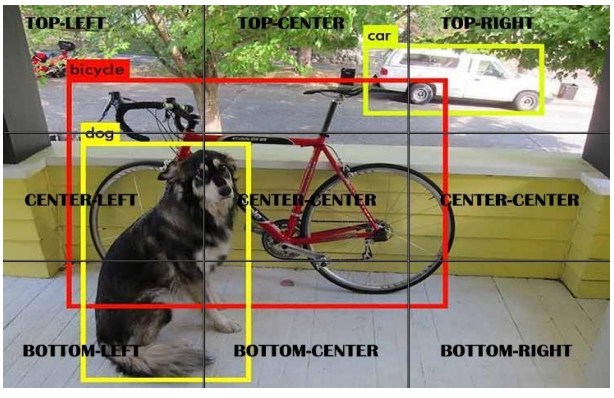


**Fig. 8: Object position estimation**

**Voice generation:**

We want to transform the categorized item into voice output to alert visually challenged people. With the assist of python, cross-platform libraries play sound and GTTS, it might be done.

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 250-258
https://publishoa.com
ISSN: 1309-3452

**RESULT:**

To test the object classifier, we can make it in real-time environment by using webcam or use the recorded video. When the webcam turned ON, the object classifier will initialize and determine the class of an object and also the position is estimated in the display, with the voice feedback.

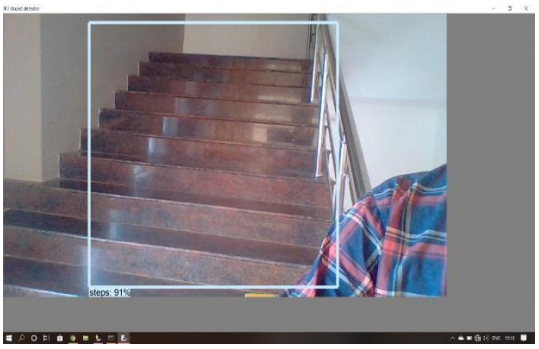


**Fig. 9: Automated system- On your center-centersteps detected**

**CONCLUSION:**

The proposed framework is a basic and effective framework that provides useful assistance and support to the blind and visually impaired. The results show that this framework is effective and extraordinary in discriminating the studies that the visually impaired may encounter. It also overcome the limitations of various frameworks identified with mobility issues affecting visually impaired people in their condition.

**FUTURE WORK:**

Future scope of this project decides to implement on Android mobile and perceive different datasets in one view with better accuracy and recognition time less than. This widening of the frame helps distinguish any type of element with a faster frame rate. The voice module is also created to modern rhythms that applicable to individuals. The model can be prepared to detect objects that the user encounters occasionally. In this way, it can be reworked for specific human needs and ensures a safer journey. Extending facial recognition to include, the application can be prepared to store data about individuals who are firmly identified with the individual, which will help separate them as friends and strangers.

**REFERENCES:**

1) H. Mao, S. Yao, T. Tang, B. Li, J. Yao, and Y. Wang, "Towards Real-Time Object Detection on Embedded Systems," in IEEE Transactions on Emerging Topics in Computing, vol. 6, no. 3, pp. 417-431, 1 July-Sept. 2018. DOI: 10.1109/TETC.2016.2593643
2) https://books.google.co.in/books?hl=en&lr=&id=ov_iBQAAQBAJ&oi=fnd&pg=PP1&dq=emotional +artificial+intelligence&ots=LBFmDXpXUc&sig =movPrEKlTAYWde5FMGus- 0WldOI#v=onepage&q=emotional%20artificial%20intelligence&f=false
3) https://www.forbes.com/sites/bernardmarr/2017/12 /15/the-next-frontier-of-artificial-intelligence-building-machines-that-read-your- emotions/#65692278647a
4) https://machinelearnings.co/the-rise-of- emotionally-intelligent-ai-fb9a814a630e
5) Kang C., Jo H. and Kim B., "A Machine-to- Machine based Intelligent Walking Assistance System for Visually Impaired Person", The Journalof KICS, Vol. 36, No. 3, 2011, pp. 195-304.
6) D. Pfeiffer, F. Erbs, and U. Franke, "Pixels, stixels, and objects," in Computer Vision–ECCV 2012. Workshops and Demonstrations Springer,2012, pp. 1–10.

7) Preetha Jagannathan, Sujatha Rajkumar,JaroslavFrnda , ParameshachariBidareDivakarachari, and PrabuSubramani"Moving Vehicle Detection and Classification Using Gaussian Mixture Model and Ensemble Deep Learning Technique"Hindawi Wireless Communications & Mobile ComputingVolume 2021, Article ID 5590894, 15 pages,https://doi.org/10.1155/2021/5590894
8) Uijlings, J.R., van de Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. IJCV(2013)
9) TensorFlow-Object-Detection-API-Tutorial-Train- Multiple-Objects- [Available: https://github.com/EdjeElectronics/TensorFlow- Object-Detection-API-Tutorial-Train-Multiple- Objects-Windows-10]
10) Real-Time Object Detection with Tensorflow Detection Model [Available:https://towardsdatascience.com/real-time-object-detectionwith-tensorflow-detection-model-e7fd20421d5d]
11) Ren, S., He, K., Girshick, R., Sun, J.: Faster R- CNN: Towards real-time object detection with region proposal networks. In: NIPS. (2015)
12) Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR.(2016)
13) Szegedy, C., Reed, S., Erhan, D., Anguelov, D.: Scalable, high-quality object detection. arXiv preprint arXiv:1412.1441 v3 (2015)
14) Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei,L.: Imagenet large scale visual recognition challenge. IJCV(2015)
15) Bell, S., Zitnick, C.L., Bala, K., Girshick, R.: Inside-outside net: Detecting objects in contextwith skip pooling and recurrent neural networks. In: CVPR. (2016)
16) Hariharan, B., Arbel a´ez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: CVPR.(2015)
17) Liu, W., Rabinovich, A., Berg, A.C.: ParseNet:Looking wider to see better. In: ILCR. (2016)13.Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Object detectors emerge in deep scene cnns. In: ICLR. (2015)
18) Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets andfully connected crfs. In: ICLR. (2015)
19) MobileNets: Efficient Convolutional Neural Netw orks for Mobile Vision Applications, Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weya nd T, Andreetto M, Adam H, arXiv:1704.04861, 2 017.
20) Object Detection Tutorial in TensorFlow: Real- Time Object Detection [Available: https://www.edureka.co/blog/tensorflow-object-detection-tutorial/]