



# SMART HELPER FOR VISUALLY IMPAIRED PERSONS

<sup>1</sup>ALETI NAGITHA REDDY, <sup>2</sup>DR M RAMA BAI

<sup>1</sup>M.TECH STUDENT, <sup>2</sup>PROFESSOR

<sup>1&2</sup>DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

<sup>1&2</sup>MGIT, HYDERABAD, TELANGANA, INDIA

<sup>1</sup>EMAIL ID: nagithar95@gmail.com

## Abstract:

Visual Impairment has been treated as an obstruction to normal functioning of vision system in human beings. The purpose of proposing this system is to aid visually impaired people because they require steady help in nearly all circumstances, particularly in the activities such as basic groceries shopping, self-navigating in indoor situations, reading magazines, newspapers, receipts, letters, bank documents etc. in order to know what is happening around them. Some of the main obstacles include the inability to move from one location to another without assistance of someone, “reading traffic signals and street signs can be extremely challenging. In order to overcome these challenges, visually impaired person might use walking cane, guide dog, and sighted guide. This paper demonstrates a generative model based on a deep neural network. The concept is accomplished using an Android mobile app that combines many technology elements. development such as voice assistant, caption generation, face detection, object detection into a single multipurpose app which helps in coordinating the visually impaired persons. The proposed app auto captures the input picture when the camera is centered through gadget and predicts results of that picture in the form of UI display as well as the audio output with confidence percentile of detection because, the algorithm is tuned to increase the probability of seeing the target depiction sentence in the training image. Using Convolutional Neural Network algorithm in the TensorFlow framework which has the necessary implementation modules of this work incorporated in it and are deployable into the application.

**Keywords:** TTS, Image caption generation, Object detection, Face recognition, Audio tutorial.

## 1. Introduction

“According to the World Health Organization, there are 285 million visually impaired people in the world, 39 million of whom are blind. These numbers are expected to rise to 550 million and 115 million by 2050. In the past, to read a newsletter, magazine, or another type of content, visually impaired people had to either find an audio version, use the braille system, or have someone read it out loud”. An individual a person who is unable to see will never experience the same emotions as someone who can see the world. This deceivability issue could be a problem that affects billions of individuals all over the planet. This research tries to overcome this problem by utilising Artificial Intelligence and Machine Learning.

Visual impairment has a significant impact on specific abilities related to visual tasks:

- i. The actions that take place on a daily basis (that requires a vision at the normal distance).
- ii. Discussion, perusing, composing (which needs an exact vision and normal separate)
- iii. Estimation of the area to be relocated and the move itself (which require a distant vision)
- iv. The tracking of an action involves increased attention to optical observation.

The need for freedom of movement in the case of visibly disabled people is recognized in today's advanced hi-tech environment. Who face fundamental societal restrictions? They are able to survive in an exciting environment without the assistance of others. Because visual information is the foundation for most tasks, those who are visually impaired are at a disadvantage because important information is lost. about their surroundings isn't available. As complete technology improves, it will be feasible to expand the aid supplied to those with visual impairments. Impairment. It's a challenging effort to naturally describe the information of

an image using properly written English sentences, but it can have a beneficial influence by standing out. People who are visually impaired find a better approach to understand the content of photographs completely. For example, this task is far more difficult than the well-studied picture categorization, which has long the computer vision community has been focusing on this topic. In actuality, a portrayal must not only capture the things in a snapshot, but also convey how these objects are related to one another and their characteristics. and the workouts they are a part of Furthermore, semantic information must be given in a certain dialect, such as English, implying that a linguistic demonstration is required in addition to visual comprehension.

## 2. Literature Review

The need for an autonomous living has been recognized in today's advanced high-tech society in the situation of visually impaired people who face a huge societal limitation. They are forced to live in unusual places with no human aid. Because visual information is the foundation for most tasks, visually impaired people are depressed since they lack the crucial knowledge about their surroundings. It is now possible to improve the help provided to people with visual impairments because to recent advancements in integrated technology. The goal of the project is to use Artificial Intelligence, Machine Learning, Image and Text Recognition to assist blind or visually impaired persons. The concept is achieved using an Android mobile application that incorporates voice assistant, image recognition, currency recognition, e-book, Chatbot, and other functions. Using voice commands and text analysis, the app can distinguish items in its environment. If you're working with a hard copy document, there is text. It is also an excellent technique for the blind to communicate with the environment using technology and to take advantage of technology's features.

Biometrics research focuses on face detection and image or video recognition. Real-time facial recognition is a fascinating field with a continually changing problem. Authentication framework for facial recognition applications. It is proposed that the PCA facial recognition system be used. Key Component Analysis is a statistical approach that falls under the broad umbrella of Factor Analysis. The goal of PCA is to minimize data storage to a large enough feature space size that data may be represented inexpensively. In the compact, PCA created a wide 1-D pixel vector with a 2-D facial picture. core parts for facial identification. of the space function. This is called the projection of self-space. The exact location is discovered by locating the Covariants matrixes own vectors, which are focused on collecting fingerprint images. By creating programming in OpenCV, Harcascade, EGeneface, Fisherface, LBPH, and Python, I constructed a camera-based real-time face recognition system and set the algorithm.

Deep learning-based object detection is a prominent application. of the technology, and it's distinguished by the robustness of feature learning and feature representation when compared to standard object recognition approaches. The study begins by outlining traditional object identification methods and explaining the relationship and

differences between scientific approaches and in-depth learning methods. It then covers the most common approaches in object detection today through in-depth learning and introduces the emergence of object identification methods based on in-depth learning. The paper's declaration of techniques focuses on the working principle of framework design and models, as well as a real-time analysis of prediction accuracy and detection accuracy. Finally, based on in-depth learning, it examines the challenges of object detection and offers various solutions for reference.

Grapheme to phoneme conversion modules are essential components of text-to-speech (TTS) systems. These modules are turned on before the phone sequence is delivered to the synthesizing procedure. Extra challenges arise when such conversion modules are employed with non-native forms of languages, such as Indian English. Because they represent American English pronunciation, many existing grapheme to phoneme dictionaries are unsuitable for use in Indian English TTS systems. As a result, this study aimed to change the current English grapheme to phoneme dictionary by integrating specific rules for one variety of Indian English, Assamese English. The proposed dictionary modification method is used in the front end of the Indian English TTS, which was built using unit selection. Considerable improvements in subjective rating are realized in both frames. After integrating the variety-specific improvements to the lexicon, the word mistake rate dropped from 46.67 percent to 7.69 percent, suggesting significant perceptual improvement.

Computer vision and natural language processing are linked by a fundamental problem in artificial intelligence: automatically describing the content of a picture. We introduce a generative model. In this work, a deep recurrent architecture is used to construct comprehensible words that characterize a picture by combining recent advances in computer vision and machine translation. The model is trained to optimize the likelihood of the goal description sentence given the training image. Experiments on several datasets show that the model is accurate and that the language it learns just from image descriptions is fluent. Our model is frequently quite correct, as we can see from our qualitative and quantitative data. For example, although the current best BLEU-1 score on the Pascal dataset is 25, our method produces a score of 59, which is comparable to human performance of roughly 69. On Flickr30k, we see The BLEU-1 score increased from 56 to 66, and the SBU score increased from 19 to 28. Finally, using the freshly released COCO dataset, which is the current state-of-the-art, we achieve a BLEU-4 of 27.7.

## 3. Materials and Methods

### Proposed system

- Object Detection:** "Object detection is a computer vision technique that allows us to identify and locate objects in an image". Using this type of identification and localization, object detection can be used to count objects in a scene, determine and track their specific locations, and precisely label them. This app can

detect then object and speak the object name in English so blind people can know the object.

- **Face Recognition:** A facial recognition system is a technology capable of matching a human face from a digital image to a database of faces. It works by locating and measuring facial features from a picture, and is commonly used to verify users through ID verification services. In Our application blind people can recognize and identify his family member's faces by focusing the camera on their faces. This app can speak the family member's names.

#### • **Image Caption generation:**

Using ML Kit, you may automatically describe Picher content is a major challenge in artificial intelligence that connects computer vision and natural language processing.

- **Color Detection:** The method of recognising the name of any color, whether it is a wall image, clothing, or other items, is known as color detection.

History of captured images & delete option which saves the captured images in history tab and also provided with option to delete.

Audio tutorial to guide the blind people.

Audio replay functionality to help the visually impaired people if any external sound disturbances are occurred constantly or occasionally while listening the audio played through this app.

Auto capture images functionality to help the visually impaired people images are captured automatically by focusing app camera for few seconds instead of capturing manually by the user.

## Implementation

The architectural design process identifies the system's components and provides a framework for subsystem control and communication. The architecture design's purpose is to construct the overall structure of the software system, as represented in the diagram below.

**Image Acquisition:** Image acquisition is the process of converting an analogue image into digital form. This usually happens in a camera or scanner, but it can be done with any device that produces analog images. Image acquisition is often used to create a digital representation of data from surveys and experiments, but it can be also used for other purposes such as printing pictures or other types of graphics

**Preprocessing of Pictures:** Preprocessing consists of decreased in duration similarly to moreover reducing the image to a defined period better to vicinity of interest specifically. It moreover complements the photo to the specified coloration choice together with is processed. Input desire of every layouts and the scale of the style via identifying the sort of requirements in every layer in particular primarily based totally surely in reality genuinely regards mostly on summation of the item of weights and biases in every layer. Likewise inside the advised CNN create the measurements of the enter picture is 64x64 in addition to the scale of the version is 45K. A fractional extra to grey variety model of the photograph does presently now not provide immoderate regular recurring performance on the similar time as contrasted to the RGB picas dealing with.

Subsequently the supported mission takes color photos even extra to be resized to 64x64 resolutions for furnished coping with.

#### Training of model:

- Model Naming: Giving a name to the model to describe the model and attach tags to the model.
- Data Type Selection: This refers to the type (Images/Text/CSV) of data that will be used to train the model.
- Data Upload: You can either upload your own data or choose from a list of public data sets. Choose from publicly available datasets such as the Celebrity Data Set (Images), the Objects Data Set (Images), the Question or Sentence Data.
- For the files (images/text files) that were uploaded and submitting to begin the upload, type category (label).
- Begin your training: Start the training by pressing the button. Mate verse's sophisticated backend will now begin processing and preparing the data you have uploaded for training.

**Classifier:** A classifier is a machine learning algorithm that organizes or categorizes data into one or more "classes" automatically. Machine learning algorithms can be used to automate tasks that were previously done by hand. They can help companies save time and money while also enhancing their efficiency.

**Output TTS:** This conversion from intent text responses to audio is known as TTS. When the user's camera is centered through gadget, input picture is auto-captured and predicts results of that picture in the form of UI display as well as the audio output with confidence percentile of detection, because the model is modified to make it more likely that the target portrayal sentence will appear in the training image. using Convolutional Neural Network algorithm in the TensorFlow framework which has the necessary implementation modules of this work incorporated in it and are deployable into the application.

#### ADVANTAGES:

- To help visually impeded persons through the Android mobile app combining the different technological development such as voice assistant, caption generation, image detection, face detection, object detection into a single multipurpose app so that they can feel the visual sense of the world through sound effects effectively.
- Building a voice enabled system that talks out the visual world for visually impaired individuals to lead the autonomous lives.
- The overall objective of this work was to develop a more accurate and accessible mobility solution for those who are blind or visually impaired to safely navigate through the urban environment.

## 5. Results and Discussions



Figure 1. Displays app home screen: This is the screen having all icons of individual applications on single screen like a brief introduction of this project.

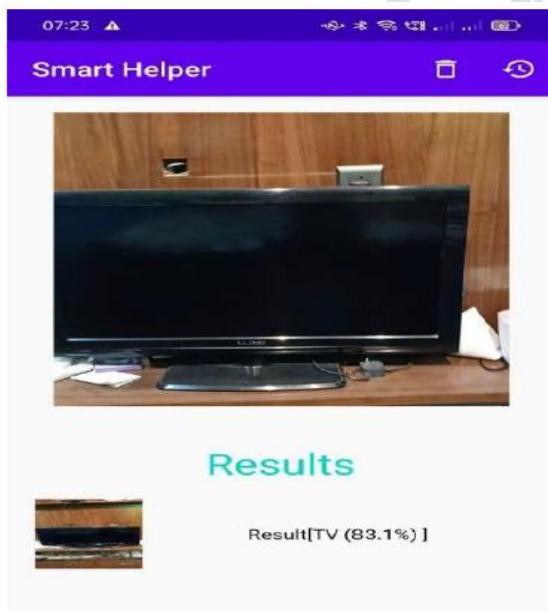


Figure 2. Object detecting capture screen: This is the captured screen with detection of an object.

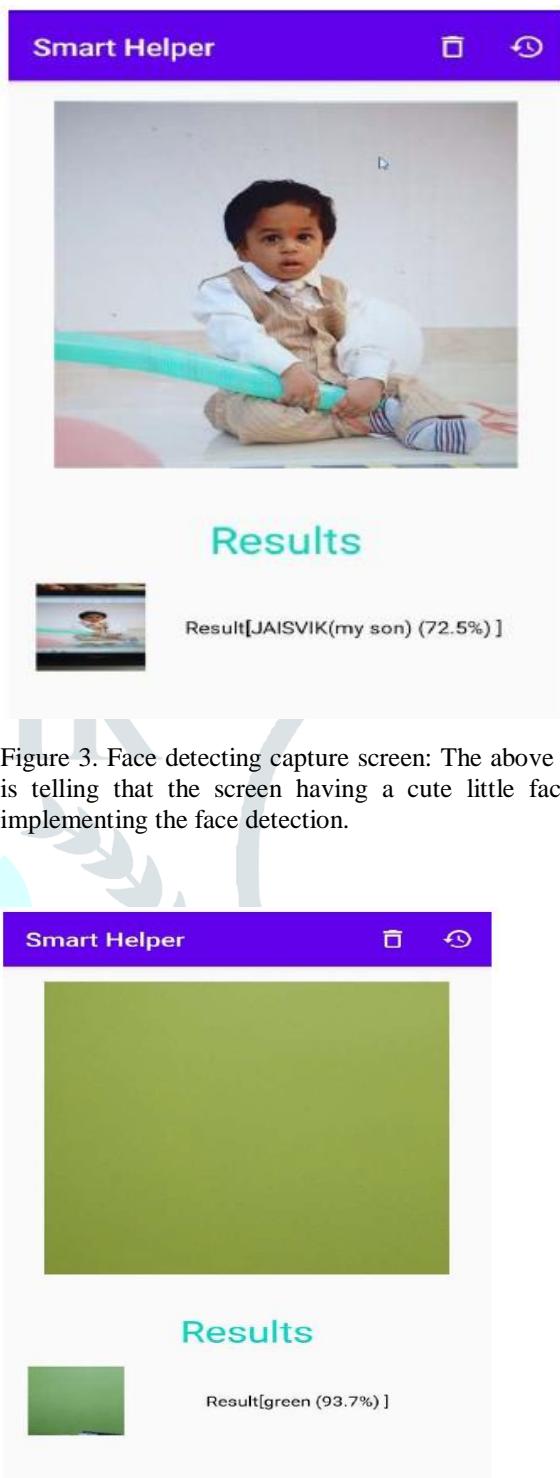


Figure 3. Face detecting capture screen: The above picture is telling that the screen having a cute little face after implementing the face detection.

Figure 4. Color detecting capture screen: This picture shows the complete captured screen of color detection means the screen having only one color.

## 6. Conclusion

The purpose of proposing this system is to aid visually impaired people who face a lot of difficulties accessing information, communicating, and mobility in almost every aspect of their daily life. For this purpose, diverse solutions have been introduced as a result of technological advancements, such as the Eye-ring project, the text recognition system, the hand gesture and facial recognition

system, and so on. However, these solutions have drawbacks such as being heavy, expensive, having less resilience, having low adoption, and so on, necessitating the development of improved ways to assist them. This work would direct the visually challenged people through the Android mobile app comprised of voice assistant, caption generation, image detection, face detection, object detection with the audio tutorial and audio repeat features as well. The app auto captures the input picture when the camera is centered through gadget and gives the detected results in the form of UI display as well as the audio output with confidence percentage of detection using Convolutional Neural Network algorithm in the TensorFlow framework which has the necessary implementation modules of this work incorporated in it and are deployable into the application. Although it's been a success in implementing the functionalities of this work, there is always a future maintenance possible in order to overcome the limitations in the current system.

## References

1. Maliha Khan, Sudeshna Chakraborty, Rani Astya, Shaveta Khepra, "Face Detection and Recognition Using OpenCV", In IEEE Conference, 2019.
2. Shubham Melvin Felix, Sumer Kumar, A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People", In IEEE conference, 2018.
3. Cong Tang, Yunsong Feng, Xing Yang, Chao Zheng, Yuanpu Zhou, "The Object Detection Based on Deep Learning", In IEEE Conference, 2017.
4. Deepshikha Mahanta, Bidisha Sharma, Priyankoo Sarmah, S.R. Mahadeva Prasanna, "Text to speech synthesis system in Indian English", In IEEE Conference, 2016.
5. Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan, "Show and tell: A neural image caption generator", In IEEE Conference, 2015.