



# Smartl

24.01.2018

---

Reuben Nellissery

Amit Jindal

Sarthak Mittal

## Overview

An android application that uses artificial intelligence to help the visually impaired understand, visualize and navigate their surroundings. Using a cascade of deep learning models hosted on a cloud server and running on the camera feed of the mobile device, the application would be able to describe the scene, estimate the distance of different objects present in the environment and also estimate the presence of free space to let the visually impaired plan their path avoiding any obstacle along the way.

## Goals

1. Describe the scene captured by the camera to a visually impaired person
2. Estimate the distance of different objects from the captured scene and provide path planning.

## Requirements

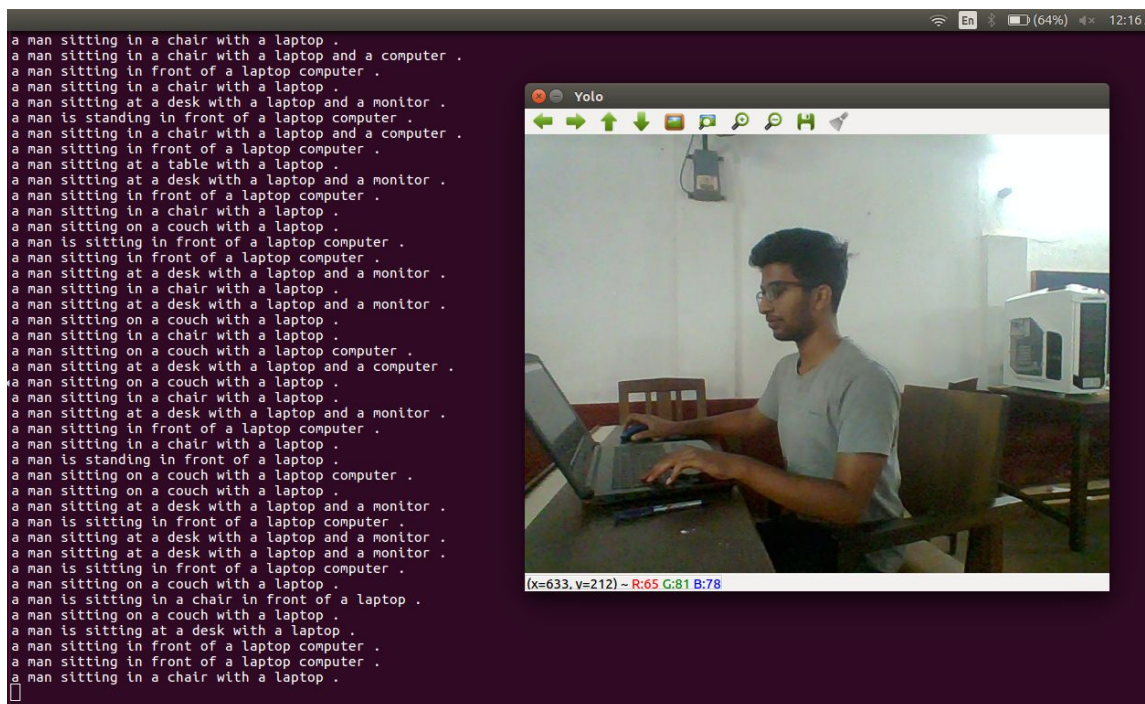
1. Android phone (Marshmallow +)
2. Decent internet connection

## Technology used

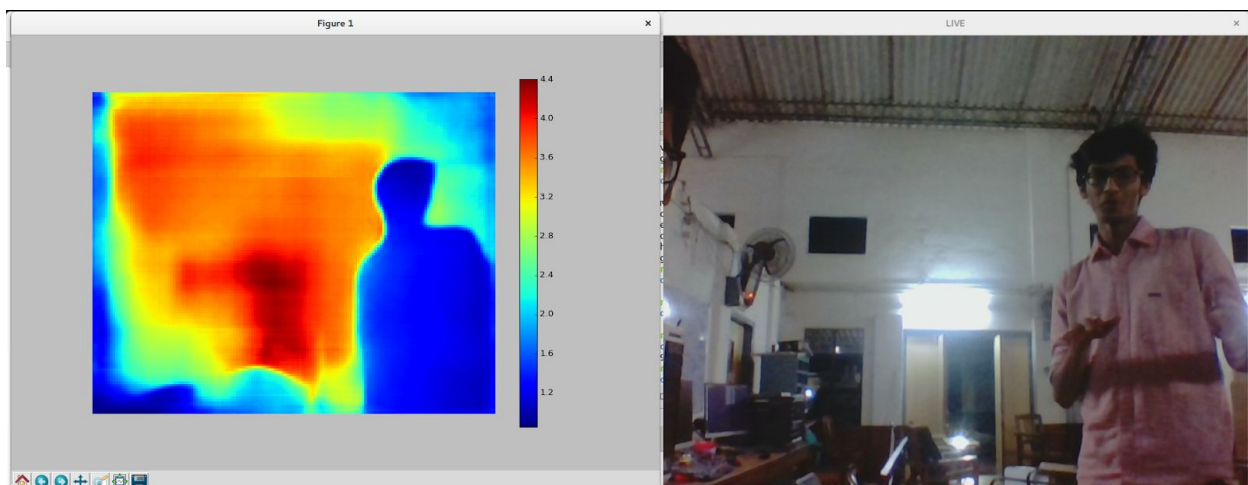
- Tensorflow
- OpenCV
- Node.js
- Express.js
- Android

## Working

- **Image Captioning:** Using a deep neural network which takes in an image from the camera feed of the android device as input and produces apt captions to describe the scene that this image portrays. Our model has a vocabulary of over 20,000 English words and has learnt to caption images from a dataset of over 150000 images. Using a text-to-speech algorithm, this description of the scene produced by the network, is converted into speech.



- Depth Prediction and Obstacle Detection:** Using a deep neural network, distance of each pixel from the camera is predicted. This depth prediction is used to find obstacles that are coming up straight ahead. Once obstacles are detected, our algorithm finds what region, left or right, has no obstacles and guides the user by outputting which region has no obstacles as speech.



- **Application:** The current version of our android application was developed with ease of use in mind. The entire screen is divided into two parts both of which act as buttons. These buttons help the user to easily toggle between the two configurations, image captioning and depth prediction. Image captioning is the default configuration for the app on startup.

## Future Plans

Our team at game\_of\_threads is constantly working to bring artificial intelligence to the masses and use this exciting technology to help the people who need it most. The future updates to SmartI would include a bunch of exciting features such as:

- Currency Detection
- More robust free space and obstacle detection for better path planning for the visually impaired
- Increasing the vocabulary size for our captioning model
- Localization and captioning of different objects of interests in the image instead of the entire scene for better understanding of the environment
- Better user-interface to make the SmartI experience more friendly