Smart-I

Made with ♥ by game_of_threads

Reuben Nellissery

Amit Jindal

Sarthak Mittal

Overview

Smart-I is 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 describes 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. Active internet connection

Technology used

- TensorFlow
- OpenCV
- Node.js
- Express.js
- Android

How to use

1. Install the provided Smart-I apk on your Android device by giving it the necessary permissions
2. On startup, the instructions on how to use the app are said out aloud to help the visually impaired

Working

- **Image Captioning**: A deep neural network is used, 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 1,50,000 images. Using a text-to-speech algorithm, this description of the scene produced by the network, is converted into speech.
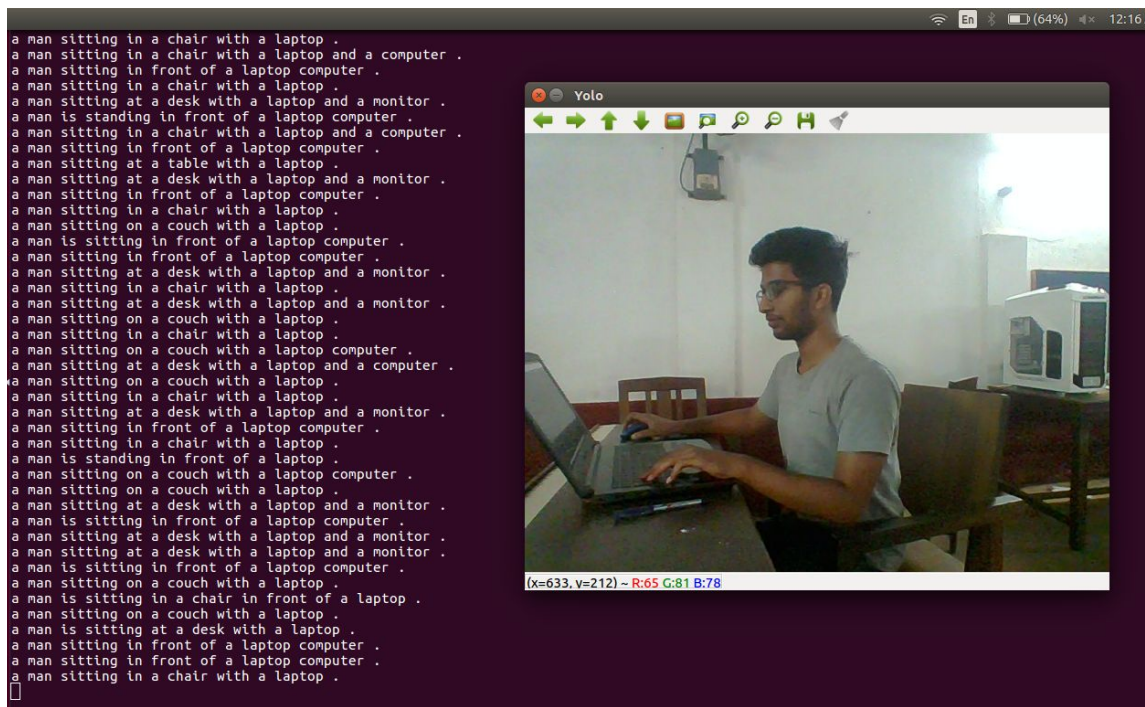
- **Depth Prediction and Obstacle Detection**: Using a deep neural network, distance of each pixel with respect to the camera is predicted. The depth prediction results are used to find obstacles that are coming up straight ahead. Once obstacles are detected, our algorithm finds what region, left or right, has no obstacles and guides the user by outputting which region to navigate towards as speech.
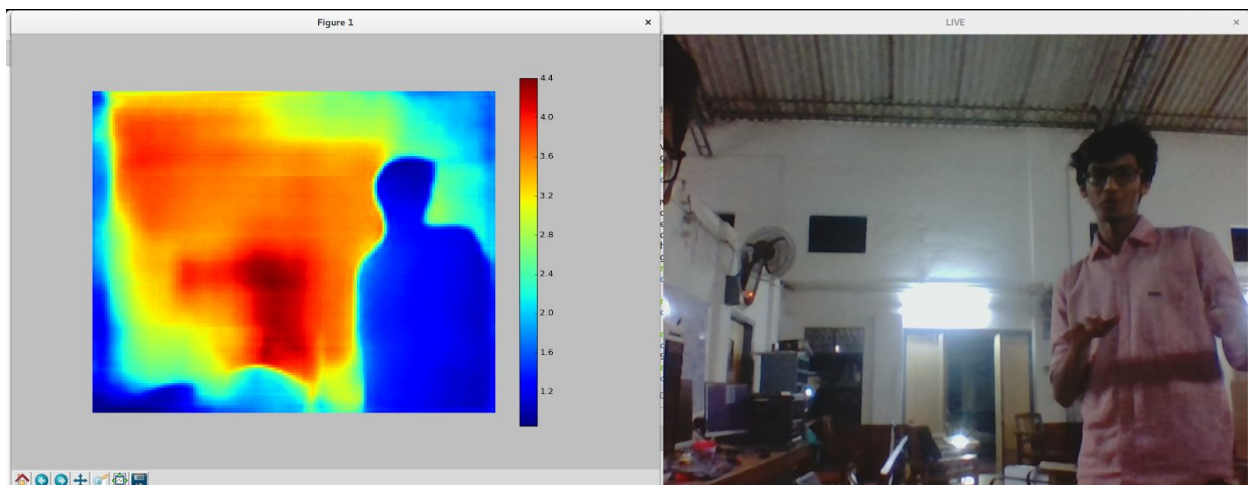
- **Application**: The current version of our android application was developed with ease of use in mind. The entire screen is divided into two parts both of which act as buttons. These buttons help the user to easily toggle between the two configurations, image captioning and depth prediction. Image captioning is the default configuration for the app on startup.
- **Server**: The app gets connected to a server that is hosted on the cloud using Azure. The app sends an image to the server and the server in turn responds with the prediction after running the image through one of our deep neural networks. NOTE: Since we are currently using the free Azure pass,virtual machines with GPU instances are not accessible and thus, we were forced to set up our server on a relatively less powerful virtual machine. As a result, the current average time of response by the server is about 10 seconds.

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
- Text-to-speech for reading documents
- 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
- Stronger and faster server running several GPU instances to increase the speed of inference