INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

AI-Powered Solution for Assisting Visually Impaired Individuals Leveraging Generative AI for Accessibility

About me

Student Name: Vijay Rawasaheb Shinde

Internship ID: IN9240777

Email ID: shindevijay595@gmail.com

Github: https://github.com/VijayShinde1996/AI_Powered_Assistant_Visually_Impaired_Individuals

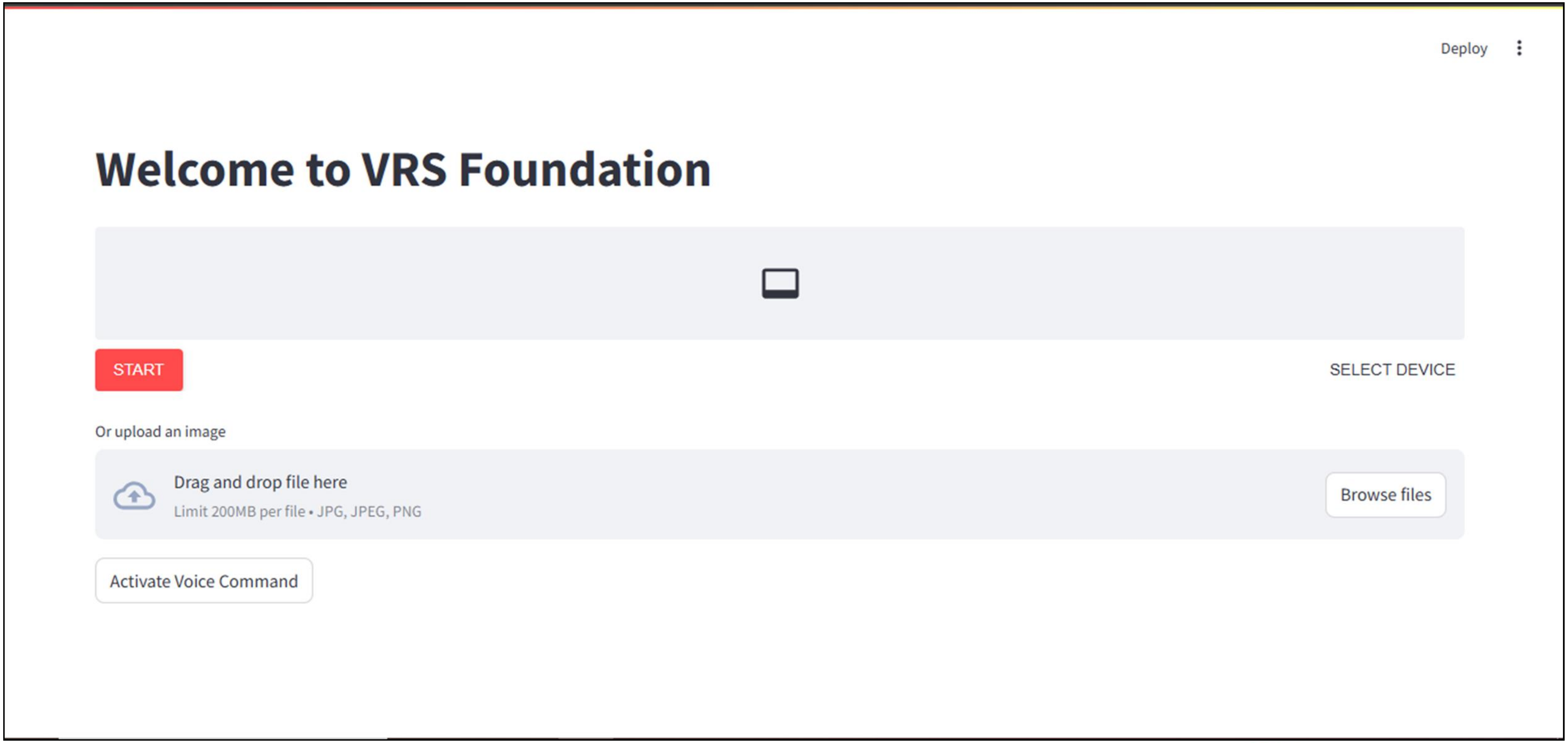
Linkedin Profile: www.linkedin.com/in/vijay-shinde-098202133

Project: Project Submission - Building AI Powered Solution for Assisting Visually Impaired Individuals

Under the guidance of: Kanav Bansal (Innomatics Research Labs)

Project Submit Date: 30th November 2024

Project Demo: Home UI



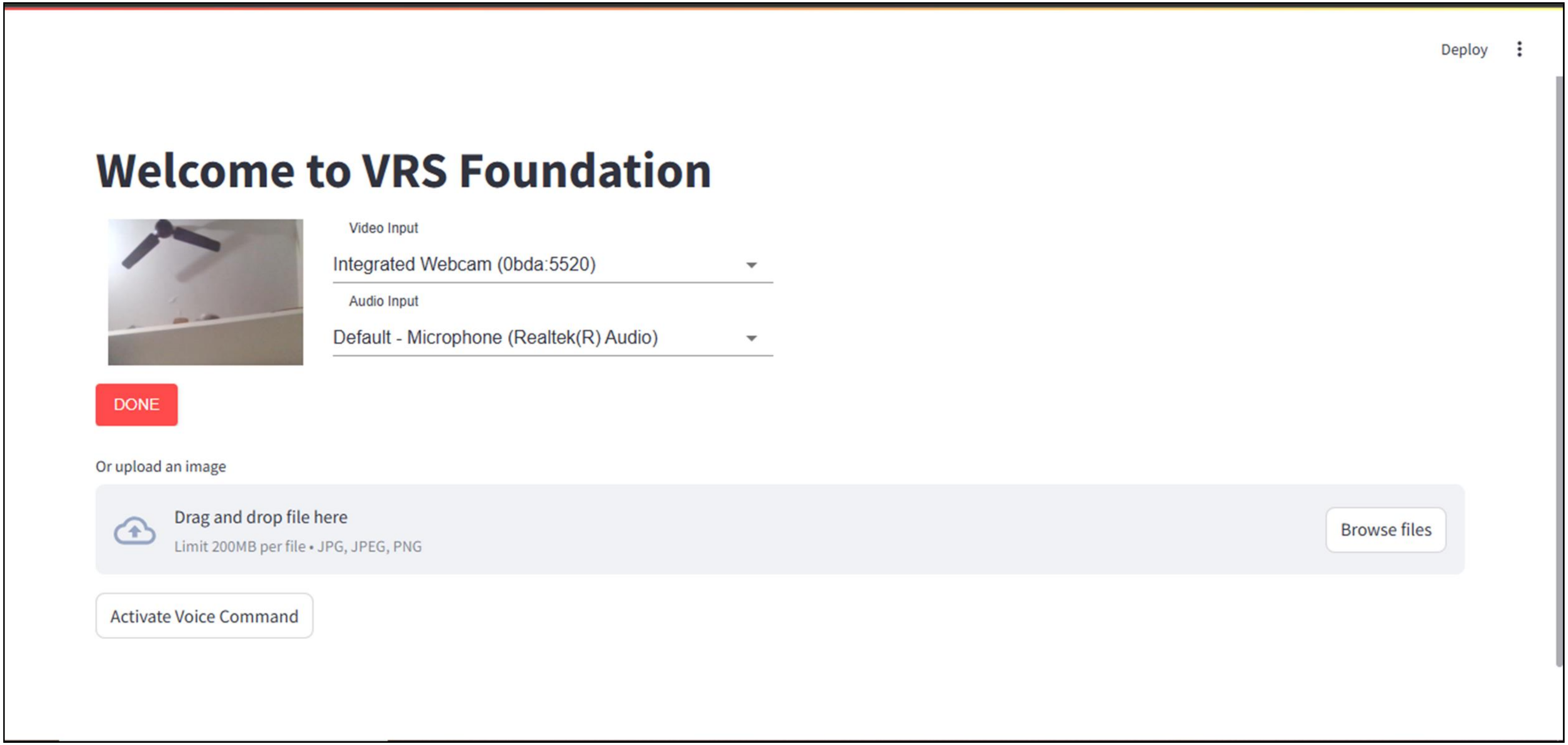
Welcome to VRS Foundation

Or upload an image

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

roject Demo: Camera & Audio Device Selection



Deploy

Welcome to VRS Foundation

Video Input

Integrated Webcam (0bda:5520)

Audio Input

Default - Microphone (Realtek(R) Audio)

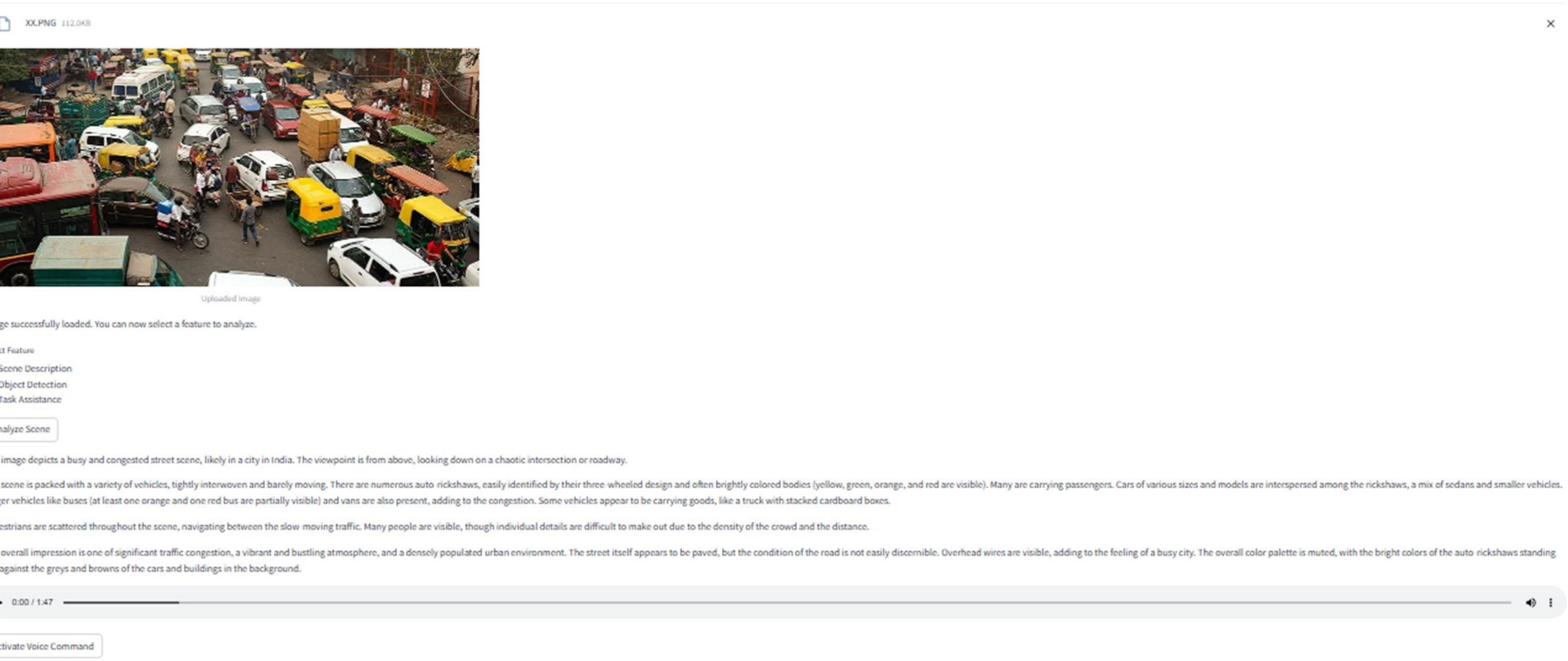
Or upload an image

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

INNOMATI
RESEARCH L

Project Demo: Analyse the Image Data



Uploaded Image

ge successfully loaded. You can now select a feature to analyze.

t Feature

- Scene Description
- Object Detection
- Task Assistance

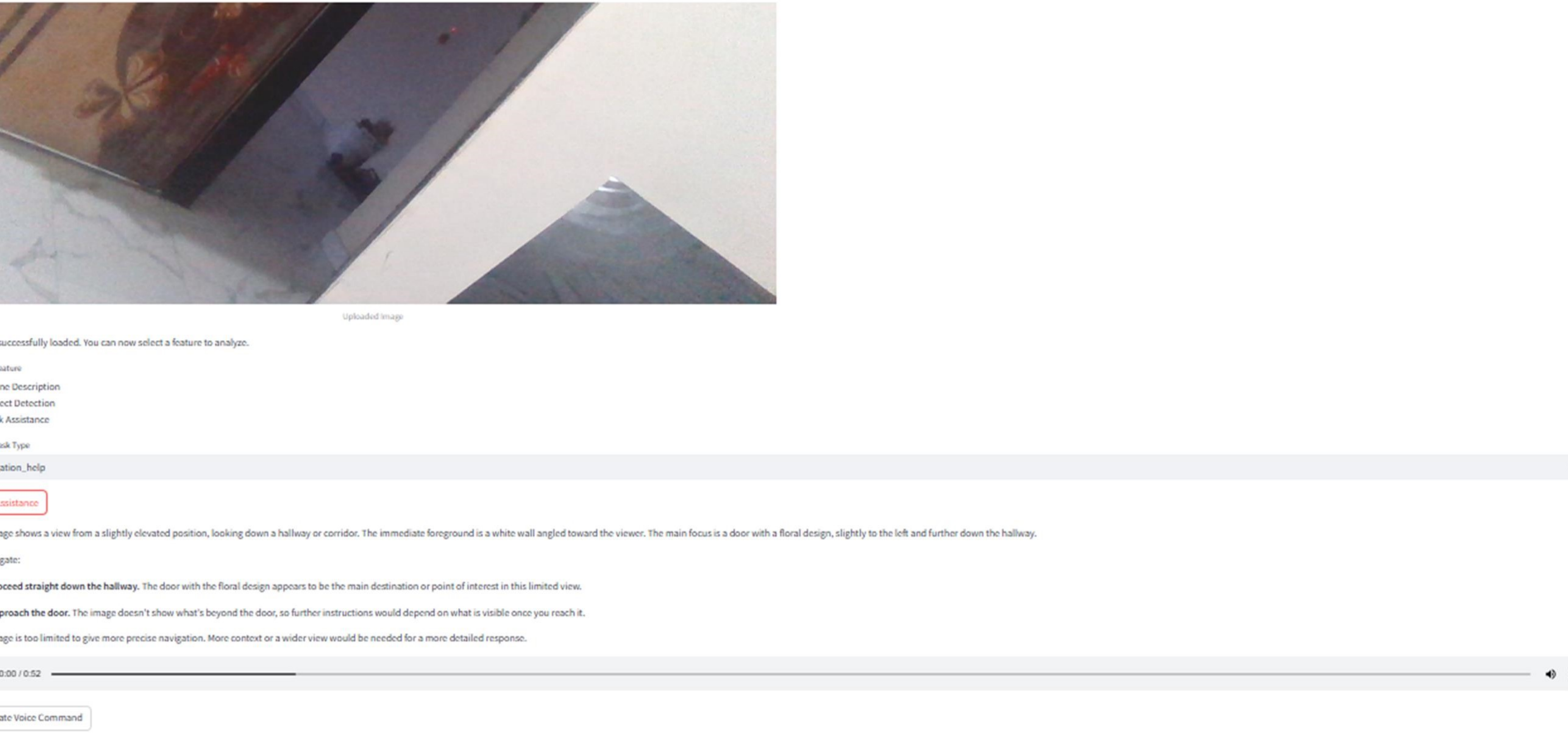
image depicts a busy and congested street scene, likely in a city in India. The viewpoint is from above, looking down on a chaotic intersection or roadway.

scene is packed with a variety of vehicles, tightly interwoven and barely moving. There are numerous auto-rickshaws, easily identified by their three-wheeled design and often brightly colored bodies (yellow, green, orange, and red are visible). Many are carrying passengers. Cars of various sizes and models are interspersed among the rickshaws, a mix of sedans and smaller vehicles. ger vehicles like buses (at least one orange and one red bus are partially visible) and vans are also present, adding to the congestion. Some vehicles appear to be carrying goods, like a truck with stacked cardboard boxes.

estrians are scattered throughout the scene, navigating between the slow-moving traffic. Many people are visible, though individual details are difficult to make out due to the density of the crowd and the distance.

overall impression is one of significant traffic congestion, a vibrant and bustling atmosphere, and a densely populated urban environment. The street itself appears to be paved, but the condition of the road is not easily discernible. Overhead wires are visible, adding to the feeling of a busy city. The overall color palette is muted, with the bright colors of the auto-rickshaws standing against the greys and browns of the cars and buildings in the background.

Project Demo: Navigate based on Camera Image



Uploaded Image

uccessfully loaded. You can now select a feature to analyze.

ature

ne Description

ect Detection

k Assistance

sk Type

ation_help

age shows a view from a slightly elevated position, looking down a hallway or corridor. The immediate foreground is a white wall angled toward the viewer. The main focus is a door with a floral design, slightly to the left and further down the hallway.

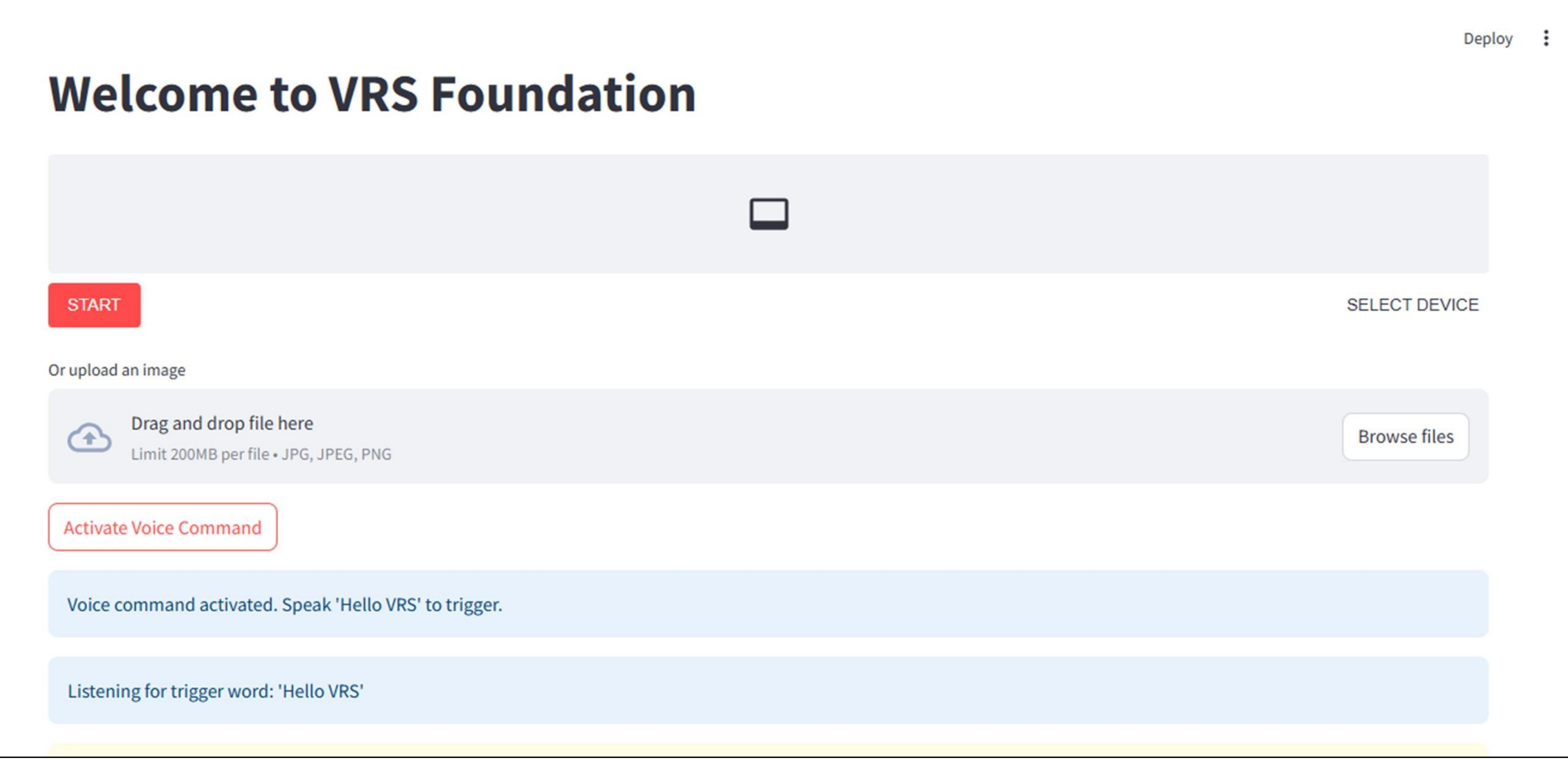
gate:

oceed straight down the hallway. The door with the floral design appears to be the main destination or point of interest in this limited view.

proach the door. The image doesn't show what's beyond the door, so further instructions would depend on what is visible once you reach it.

age is too limited to give more precise navigation. More context or a wider view would be needed for a more detailed response.

roject Demo: Voice Assistance



Welcome to VRS Foundation

Or upload an image

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

Voice command activated. Speak 'Hello VRS' to trigger.

Listening for trigger word: 'Hello VRS'

roblem Statement

Visually impaired individuals face significant challenges in perceiving and interacting with their environment. Everyday tasks that rely on sight, such as recognizing objects, reading text, and understanding surroundings, become barriers. Therefore, there is an urgent need for an intelligent, adaptable, and user-friendly solution that can offer:

- **Real-time scene understanding:** Interpretation of surroundings and scene context for better awareness.
- **Text-to-speech conversion:** Reading visual content to enhance accessibility.
- **Object and obstacle detection:** Helping individuals safely navigate by identifying obstacles and objects.
- **Personalized task assistance:** Providing support for daily activities, such as identifying items and reading labels.
- This project seeks to build an AI-powered application that offers these functionalities through a seamless interface, enabling visually impaired users to interact with the world more effectively.

rocess Flow

- **Opening the App:** The user is greeted with a welcome message, "Welcome to VRS Foundation.“
- **Capturing Image:** The camera opens automatically, and the user is prompted to take a photo.
- **Selecting Action:** After capturing the image, the user is asked what they would like to do with the image. Available options include:Scene Description
  - Object Detection
  - Task Assistance
- **Providing Results:** The app processes the image based on the selected action and provides either an audio description or guidance.**Voice Command Activation:** The app can also be triggered by saying "Hello VRS," allowing the user to perform the same tasks without touching the interface.
- **Voice Command Activation:** The app can also be triggered by saying "Hello VRS," allowing the user to perform the same tasks without touching the interface.

asks & Features

The app is designed to offer the following core functionalities using AI and Streamlit:

- **Real-Time Scene Understanding:** The app generates a textual description of the scene in the uploaded or captured image, helping users understand their surroundings.
- **Text-to-Speech Conversion for Visual Content:** Extracts text from images using OCR techniques and converts it into audible speech, allowing users to hear any visual text or labels present.
- **Object and Obstacle Detection:** Identifies objects and obstacles within the image, helping visually impaired users navigate safely.
- **Personalized Assistance for Daily Tasks:** Provides task-specific guidance based on the image, such as recognizing objects, reading labels, or giving directions for daily tasks.

ode Walkthrough

Import Libraries:

The necessary libraries are imported, including **Streamlit** for the web interface, **Pillow** for image processing, **gTTS** for text-to-speech conversion, **Google Vision** for image analysis, and **Langchain** for using generative AI models.

API Initialization:

API keys are set for accessing Google’s Generative AI and Vision models.

Langchain’s ChatGoogleGenerativeAI is initialized to process image analysis requests.

Error Handling:

A function is defined to handle and log any errors that occur during the execution.

Image Analysis:

The function converts the uploaded image into a byte stream and sends it to the Google Generative AI model to process and return content like scene descriptions, object detection, or task-specific information.

Text-to-Speech Conversion:

The extracted text (like scene descriptions or object details) is converted into speech using the **gTTS** library and returned as an audio file.

Webcam Capture:

A class is defined to use **Streamlit WebRTC** for capturing an image from the webcam. The first captured frame is saved for analysis.

Voice Command Listener:

The app listens for the voice command "Hello VRS" using **SpeechRecognition**, which triggers the app’s main functionality.

Main App Logic:

The app greets the user and displays options to either capture an image using the webcam or upload one manually.

After selecting an image, the user can choose from features such as **Scene Description**, **Object Detection**, or **Task Assistance**.

The app then analyzes the image based on the selected feature and provides both text and audio output.

Voice commands are continuously monitored to allow hands-free interaction.

onclusion

- This AI-powered assistive application, developed using **Streamlit**, **Google Generative AI**, **Langchain**, and **Text-to-Speech**, provides valuable functionalities to visually impaired users. By offering scene descriptions, object detection, text reading, and task assistance, this app empowers users to interact more effectively with their environment and navigate safely. Voice activation further enhances user interaction, creating a seamless experience.
- With potential future improvements like personalized assistance and multi-language support, this app could become a powerful tool for improving the independence and quality of life for visually impaired individuals.

uture Scope

The current app offers a robust solution, but there are several areas for improvement and new features that could be added to enhance the experience for visually impaired users:

- **Personalization**:
  - Introduce user profiles to save preferences, such as preferred speech speed or tone.
  - Customizable settings for object detection to focus on specific items like food or common household objects.
- **Real-time Navigation**:
  - Implement real-time navigation assistance using a mobile phone camera to detect obstacles and guide the user through their environment.
  - Integrate with a GPS system for location-based assistance (e.g., guiding through a building or outdoor spaces).
- **Multi-language Support**:
  - Add multi-language capabilities, allowing users to interact in their native language.
- **Enhanced Object Detection**:
  - Improve object detection by training custom models for specific use cases, like identifying medical equipment, currency notes, or reading labels in different fonts.
- **Integration with Smart Devices

THANK YOU