Writing 1 - Jack

If you are visually impaired, there may be times when you need to navigate the world on your own. We are developing a mobile application to help people navigate on foot safely in potentially hazardous everyday environments. Our application is composed of a direction-giving service built on Google Maps, but overhauled to add features and improved usability for the visually impaired.

With this application, the user can verbally request gps on-foot directions. The user holds up the phone camera to face in front of them while walking, and the app detects objects seen by the camera, such as cars, cross-walks, street signs, roads, bikes, etc and speaks them out loud to the user. Additionally, if a street sign is detected, like a stop sign, railroad crossing sign, construction warning sign, etc, the app can identify the type of sign it sees. With these functionalities, this application can help a visually impaired person more safely walk from point A to point B.

A practical example of this application can be seen in the case of a blind user crossing a road. Crossing a road is dangerous and can be scary for a blind user - it can be hard to discern the proper place and timing to cross. The application helps alleviate uncertainties by notifying the user when a crosswalk is seen by the camera. With the aid of this application, the user can gain a complete understanding of his surroundings before stepping onto the street. Moreover, if the blind user is concerned about oncoming low-sound vehicles, he can effectively “look both ways” before crossing the road by facing his camera left and right.

The problem of a computer program accurately identifying objects in the real world using a camera feed is a difficult technical challenge that may not have been feasible just a few years ago. In order for the application to detect and identify objects seen in the camera, we make use of a powerful technology called machine learning. Machine learning is a sub-area of artificial intelligence that involves computer programs using data to improve automatically through experience. For example, a machine learning application can become good at classifying cats versus dogs after being given thousands of pictures of cats and dogs to learn with.

Utilizing machine learning requires access to large amounts of data, and it can be difficult to find or create the necessary data for a machine learning task. We ‘teach’ our application by repeatedly exposing it to volumes of images so that it can become accurate at detecting objects in the novel, real-world context. We will approach the challenge of gathering sufficient data by experimenting with relevant datasets that are publicly available for use from databases like ImageNet or data-hosting websites like Kaggle.com.

Compared to similar applications and products directed towards a target audience with sight problems, our application has characteristics that give it an independent purpose. Products like the MyEye2 electronic glasses are more cumbersome and require additional hardware beside a cell phone. Additionally, our application requires minimal cell connection for gps navigation, while other mobile applications like Be My Eyes heavily require internet or cell reception to accomplish what our application does with machine learning.

Altogether, our application does more than just visual machine learning. The machine learning functionalities are integrated with speech recognition technology and map navigation, which pre-exist as reliable software. With our team of software engineers, we can combine these features into a user-friendly, convenient mobile application. The end result is a powerful assistive application that simulates the job of a seeing eye dog for the visually impaired.

Writing 1 - Rick

Worldwide, many people are visually impaired; that is, unable to have their eyesight corrected by glasses or contact lenses. While public spaces sometimes have accessibility features to assist with this difficulty (for example, a crosswalk signal that makes a chirping sound to alert pedestrians that it is safe to cross), these assistive features are by no means universal and are not always well-maintained. An alternative solution is for visually impaired individuals to provide their own accessibility aids, like a white cane or guide dog. While these have proven useful in many respects, these tools also have limitations and issues. Our team notes that some of these limitations are especially important in a world that is becoming increasingly technical. For example, guide dogs cannot read street signs or use color as a cue.

Here, the tech industry can provide unique solutions which are more feature-rich and less cumbersome than older methods. In researching this question, we found many existing tools. Equipment such as eSight glasses or MyEye 2 offer an impressive selection of utilities. These tools typically involve a wearable camera affixed to glasses which can connect to a smartphone using wireless technology. All this convenience comes at a price, however: smart assistive eyewear typically costs multiple thousands of dollars, which is oppressively steep for many people. A deeper problem is that these devices require the user to have some sight, making them less helpful for people who are nearly or completely blind.

Our proposed solution is designed to be cheap and widely available: a smartphone app. The app's interface will be well-integrated with pre-existing accessibility features built into major mobile operating systems. For instance, both iOS and Android have screen reader capabilities. For users who can see the screen, the interface will be clean with large symbols and lettering. Users can interact with the app using voice commands, or by directly interacting with the screen. Thus, we do not anticipate the platform to be a barrier for more severe levels of visual impairment.

The functionality of the app is meant to cover safety and utility, areas that existing solutions have not focused as much on. Specifically, our app will use the phone's GPS to guide the user along a route (similar to how a guide dog would). It will also use the phone camera to provide information on possible obstacles along the route, as well as to identify information sources such as street signs to the user. All this information will be relayed using voice cues from the smartphone, and the user will be able to customize which ones they want to hear so that our app can be usable for anyone, regardless of their personal situation. Factors such as variation in phone position or camera quality make this task a difficult one.

To deal with this, our app will employ machine learning. In a nutshell, machine learning is a process in which a computer is trained to put labels on pieces of data (such as photos or text); it "learns" to categorize things. Now that huge repositories of photos are publicly available online, machine learning is a state-of-the-art solution for image recognition. Modern smartphones are incredibly powerful, and we intend to take full advantage of this. Certain types of machine learning algorithms are excellent at identifying objects in an image. Our team will be training one to find street signs in an image so that the user can receive information about their surroundings that is relevant for their personal safety. Further machine learning solutions will be employed to find important objects in images (such as traffic lights). Thanks to publicly available datasets, we will not need to spend valuable time and resources to personally gather relevant images. Initial investment in our project will go directly to development of the app.

Writing 1 - Selin

In a world built to accommodate those who are neurotypical and able-bodied, the limiting factors for those who have certain disabilities - such as low or non-sightedness - are strikingly apparent outside of spaces that are specifically made to be accessible for them. While things like braille menus and signs are becoming more commonplace, there are still few accessible solutions for people who are non-sighted to be able to navigate an urban world autonomously. We aim to design a tool using current technologies to offer a new means of understanding the environment to those people so they can live their lives with less limiting factors. To that end, my associates and I have taken on the responsibility of designing a product that we hope will be part of redefining the paradigm of accessibility devices as we know it.

The working title for this project is Project Digital Orientation Guide (D.O.G.), and in summation, we are building an Android application that uses a GPS in conjunction with image classification models that are able to recognize objects from images to simulate the functions of a seeing eye dog. Through this, we will be able to teach a program to recognize and decipher common visual street signs and cues, much like a seeing eye dog would, and translate those commands into audible cues that the user can easily make sense of. This project will include object detection and collision detection modules that will use a phone camera to recognize objects in the area, a communications module to relay information to the user, and mobile application to run everything together, as well as provide extra safety checks via GPS tracking to ensure that the safety of the user is not being compromised.

The biggest portion of the project will consist of the first two modules we've mentioned: object and collision detection. At a high level, we will be developing an algorithm that can detect basic objects like cars, crosswalks, road blocks, and street signs, and send that information to the communications protocol so it can be relayed to the user. The object detection portion of our application will be dedicated to detecting things like traffic lights, stop signs, crosswalks, and pedestrian walking signs. In summation, this module will be in charge of processing information. Our collision detection, however, will be responsible for detecting things such as objects obstructing the path in front of the user, manholes, cars, and other objects that can physically obstruct the user. In both cases, the names and relative positions of these objects will be recorded and sent to the communications protocol.

The next cornerstone of our project will be developing a communication system that will be able to recognize basic commands from the user, and be able to read out information sent from the object collision and detection modules. This will most noticeably be in charge of text-to-speech interactions, but it will also be in charge of accepting and processing basic commands that the user prompts, such as repeating information, inputting an address to the GPS, asking to check for specific objects, etc. This module handles most of the user interface related to the application.

Finally, the entire brain of our operation, our Android based mobile application. The application is in charge of hosting our software and integrating the mobile device's hardware with the commands that are requested by the other modules - being the camera for object detection and the microphone and speakers for communications. The application will also feature a GPS service that will serve as a double-check to make sure that the input generated by the object detection module make sense for the geographical coordinates of the user, i.e. if the user is near an intersection, the application should be detecting a traffic light or stop sign.

Our vision is to create an affordable, safe, and effective means of making the world a more even playing field. We hope that this new means of perceiving the environment will instill a new confidence in our users to be able to navigate the world independently. With all of the advanced technologies our product utilizes, it undoubtedly has the potential to be something great, but more importantly, it's going to be making a difference in the world. Since we often underestimate the power of having options available to us in the market, even if this product unexpectedly falls into obscurity, we will have at least given a new option to those who already have very little to choose from.

Corrected Writing 2 - Comparison to Other Products (Rick)

The idea of using modern technology to provide assistance to visually impaired people is by no means a new concept. Many products exist; most are wearable devices which provide an augmented reality (AR) experience. Perhaps the most feature-rich example of such a device is the Orcam MyEye. This is a camera which mounts to eyeglass frames using magnets. The camera connects to a smartphone app via Bluetooth, and offers features such as text reading or facial recognition for frequent contacts. Conceptually, it is meant to assist with day-to-day life. There are many variations of this product (e.g. NuEyes, eSight, Acesight), but all are targeted at similar use cases. Our product is meant to assist the visually impaired in an area that these products lack: navigation. As this is such a central part of modern life, older solutions exist (e.g. white canes or guide dogs), but these have their own limitations. Our app provides an elegant, high-tech solution. Aside from having unique functionalities from the other products, it has several other advantages which set it apart.

Our product is cheap. The tools used to make it (i.e. Android app libraries, machine learning libraries) are free and open source, and the data used to create the artificial intelligence powering it is also publicly available online. With this low cost of creation, we expect to be able to offer the app for free. Compared to other visual assistance products, this removes a significant barrier: the Orcam MyEye, for example, currently costs around $3,500. Alternatives which we previously mentioned are similarly expensive. Unfortunately, the app will not immediately be available on iOS, but this is justifiable because news sources estimate that Android holds approximately 70-80% market share worldwide. Additionally, our app is built to be usable by people who are almost or totally blind. Many of the AR solutions and camera attachments we previously mentioned require at least partial sight to operate and setup. Our app's ability to recognize voice commands allow it to be used without needing to see the screen.

We designed our app to be a unique and accessible solution for the visually impaired. Its targeted feature set and low price allow it to be a useful tool for anyone who has difficulty seeing. It could be used in conjunction with a more expensive product, or completely independently. We expect the usefulness and accessibility of our app to set it apart from other products in this field and make it appealing to many people.

Corrected Writing 2 - Societal and Global Impact (Jack)

The broader societal need we are seeking to address is the need for an accessible tool which provides safe on-foot travel for the visually impaired, especially in urban areas. Our product is accessible because it is a mobile application, so the user does not need to pay for additional expensive hardware.

If used widely, it could positively impact society in several ways. First of all, our application could reduce the number of accidents involving visually impaired people. It can aid in emergency situations like when a blind person is isolated from the help of others. The capacity to prevent accidents can be seen especially in urban settings, which the application is specialized for. Additionally, sidewalks, crosswalks, and interiors like grocery stores are examples of spaces that can pose an accessibility problem to visually impaired people. This application can help users navigate these types of spaces more effectively by alerting them of obstructions like people, cars, or stairs and providing details like the type of traffic sign or the color of a traffic light. In this sense, the application would grant visually impaired people comfort and a degree of independence.

This application will not require regulation. It is data-safe. In other words, since the app does not collect or store any information about the user, there is no risk of exposing sensitive data. Additionally, the application has minimal reliance on a network connection which is only used to support the direction service; so the user never needs to connect to a wifi network. Since this is simply a tool to aid the visually impaired, and the use of this application does not impose upon those around the user in any way, no outside regulations are necessary.

One concern may be about the possibility of someone purposely sabotaging the object detection to detect the wrong objects, or fail to detect the right objects. In fact, this is not a realistic concern because the object detection operates based on a pre-trained model that is handcrafted by the developer. Simple checks can be performed in the code to ensure that the correct model is being used by the application .

This application can, however, be put to bad use by the user himself. The issue is over-reliance. This application can help in emergency situations and in low-danger environments, but it is not advised for a severely visually impaired user to choose this application over the physical aid of another human or guide dog if those options are available - especially in high vehicular traffic areas. To reinforce this notion, the application can warn the user when in or near a high-traffic area like a city or highway.

This leads to some international-use concerns related to our application. First of all is the topic of liability. If over-reliance on the application leads to injury, the issue of liability arises. We can include a disclaimer upon the first-time opening of the application that affirms the app's intended use, and warns against over-reliance on the application in hazardous environments. If necessary, we can also include an agreement that places liability of injury on the user, rather than the developer. A separate concern is the variance in the precision of Google Maps across the globe, given that Google Maps powers our direction functionality. If the user is in a location known to be un-mapped or sparsely-mapped, we can warn the user of the potential direction-giving inaccuracies .

Corrected Writing 2 - Target Audience (Selin)

The target demographic for our application has a large range of intended users; the app is helpful for a completely blind user, as well as users with mild-moderate vision loss. Since there is no existing aid on the market that was built for the completely blind in mind, we specifically designed this app to relay information in a way that completely blind people can best understand it, which is also inclusive of those who are mild-moderately visually impaired people who have troubles seeing close-up and far-away details.

There is a realistic potential for this tool to be used widely because of the pre-existing pervasiveness of cell phones. By publishing this application on the Android app store, we are able to avoid having to pay any publishing or maintenance costs, ensuring that anyone with an

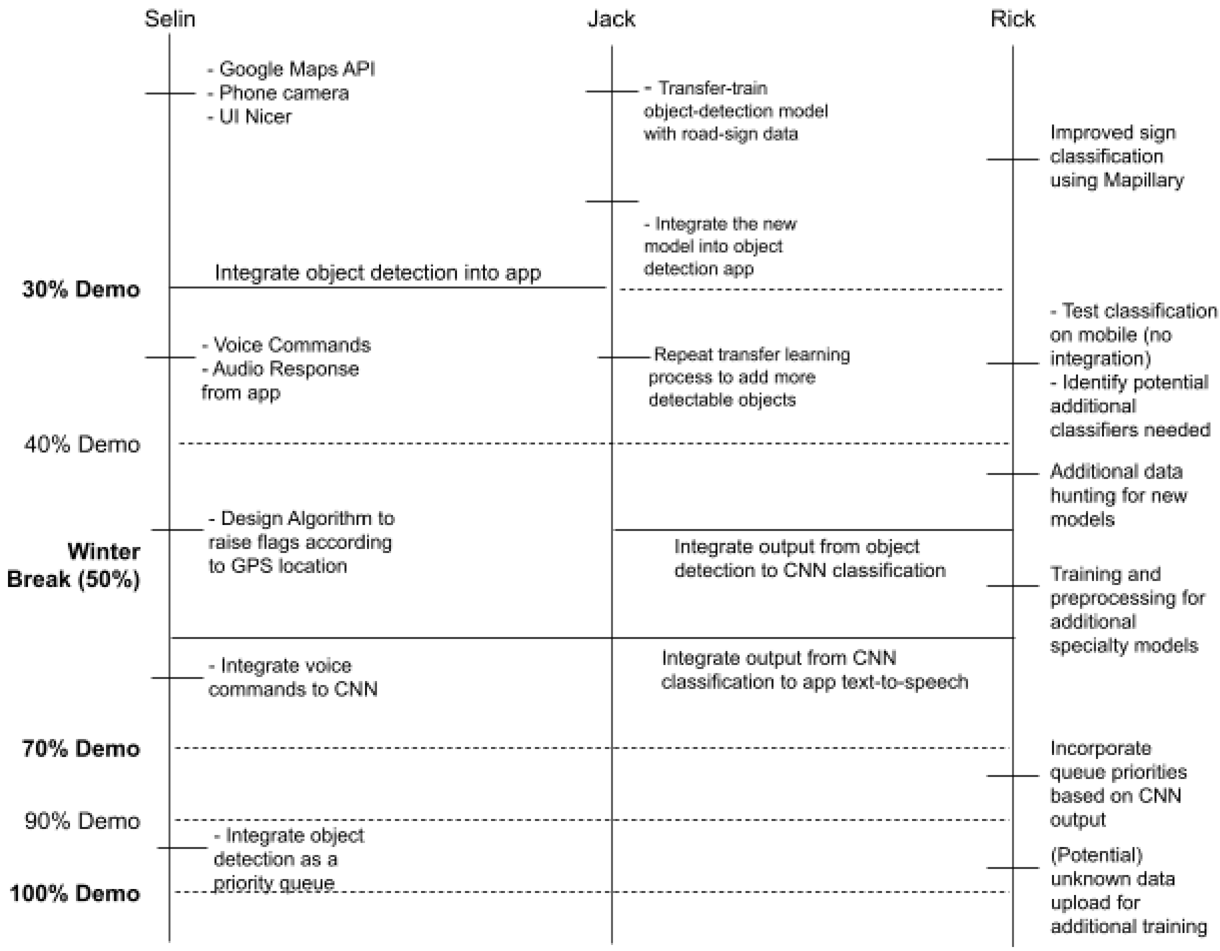
Android smartphone will have access to this tool without us needing to charge for service in any way.

Additionally, many people in the US have access to an Android device, so there's the benefit of having no extra costs to access the software, with the additional benefit that the user would not have to acclimate to using new, bulky, and unfamiliar gadgets.

Through working closely with the American Foundation for the Blind, an existing and well-known organization that focuses on improving the accessibility of everyday activities, we will gain an invaluable means of certifying our application's reliability, and along with it, countless new contacts through which we can market our product.

Writing 3

Timeline Diagram (Full Team)



Technical Innovation (Rick)

Project D.O.G. incorporates several state-of-the-art machine learning technologies and industry-standard Android app libraries. Our app is a Kotlin-based Android app using the Google Maps API and text-to-speech API. Its machine learning capabilities are powered by multiple types of model architectures; our final product will run object detection using a fine tuned MobileNet model, and then get more specific results using a multilayer convolutional neural network (CNN). In other words, the app camera will do some basic object identification (e.g. find a sign in the frame) using MobileNet and then use the CNN to get more useful results before reporting them to the user (e.g. if MobileNet detects a sign, the app will then use a CNN to determine *which* sign it is).

As described in previous sections, there are many existing products which seek to provide a technology-based visual aid for people with a high level of visual impairment. Unfortunately, these products are not open-source, so it is impossible to say precisely what technologies they run on. In general terms, our product is novel because it offers this visual aid service to more people at a low price. A smart camera such as the MyEye OrCam likely uses CNNs for image classification (these are state-of-the-art), but our layered solution using MobileNet alongside CNNs is likely a novel approach. D.O.G. is designed to be usable by someone who is fully blind; other apps which do this use crowdsourcing to identify what's on the phone camera, while D.O.G. uses machine learning. Thus, the technical innovation of D.O.G. is a crossroads of machine learning technology with cheap and accessible hardware requirements.

Technical Feasibility (Jack)

There are several existing tools and technologies that we use to build our project. First of all, we use Tensorflow to implement the machine learning parts of our project. Tensorflow is a free and open source software library for machine learning. On the Android device itself, we make use of a variant of Tensorflow called Tensorflow Lite, which allows developers to run TensorFlow models on mobile, embedded, and IoT devices. Since Tensorflow Lite is designed to run on mobile, we are confident that we can use this tool to perform object detection and object classification in real-time. Additionally, Tensorflow provides a selection of sample Android apps which we tested to confirm that machine learning can be done on a wide range of Android devices with acceptable latency.

Another tool we are using is the Google Maps software development kit for Android. Using this SDK, we can make customized, static and dynamic maps, Street View imagery, and 360° views. To implement audio functionality such as user voice commands and audio narration, we make use of two tools: Android Speech Recognizer and Android TextToSpeech. These are classes provided by Android that we can use when writing the code for our app. Since these official tools are readily available, we do not have to implement natural language processing from scratch. In summary, the tools that we are using to build our project are feasible, and we are confident in these tools' ability to support our development process.

Cost, Risk, and Risk Mitigation (Selin)

Costs for this project are minimal for two main reasons, the first being that the software we are using to develop our application is free to use and does not require particularly powerful computers to run - and the second being publishing applications on the Android store is also free. It is in part because of our materials being free that we will be able to make the application free and without advertisements for all users. Development costs for a potential new employee/partner would come from having to buy a new computer to code on and/or an Android device to run the application.

At this point, it is hard to estimate the length of the project, but our initial estimate is about 10,000 lines of code, as is the average for a mobile application. This figure includes the mobile application and the training module that we will be writing that will not be a part of the application itself.

There are four main milestones that we are aiming to reach, which are bolded in the timeline diagram above. Our first milestone will be on November 24th, 2020, which will be our 30% implementation of the project; at this point we will have a mobile application that is fully navigable and that can perform object detection. Our second milestone will be on January 1st, 2021, where we will have a 50% implementation completed; this implementation will add a more refined object detection process, and will only detect objects related to pedestrians, like signs and crosswalks, and it will also include voice commands and audio cues. Our third milestone will be our 70% implementation on February 9th, which will be our working beta version of our application; after this milestone, the app will only require mild improvements to improve functionality and efficiency. Our last milestone will be the delivery of our final app, which will be completed on April 6th.