Assistive Technology to Recognize Objects Through A Camera

William Go, Computer Science

Mentor: Dr. David Johannsen

Kellogg Honors College Capstone Project for RSCA 2024

Abstract

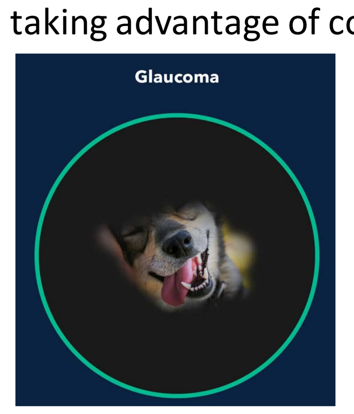
Many aspects of life prove to be hostile for the blind or visually impaired yet through innovation and technology, it is possible to gain more independence. This project aims to bridge the gap between the visually impaired and their surroundings to enable them the ability to navigate and interact with objects more effectively. Using object detection technology and integrating it into a real-time application with auditory feedback, it is possible to gain further autonomy and awareness of one's surroundings. This application uses TensorFlow's object detection model for computer vision to relay information based on what is seen within a camera. The design includes visual camera feedback for those with low vision and auditory feedback that regularly scans and repeats object information. Custom detections and further development of the library of detectable objects are possible through training TensorFlow's model. Deploying this technology as an application allows for mobility and flexibility to be accessible in any scenario within daily life whilst being more financially accessible compared to current technologies in the market. Creating assistive technology is integral to inclusivity and accessibility from within the home to the outside world.

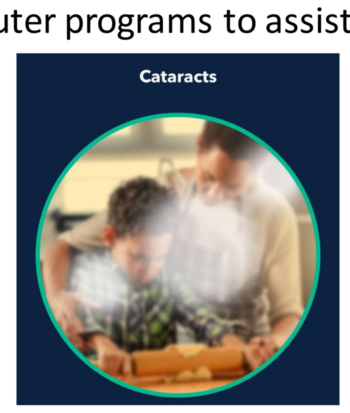
Objective

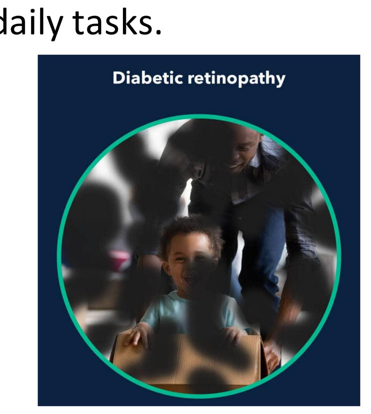
Many visually impaired individuals require assistance to act on many daily tasks. Using tools and applications are a way to become more independent and capable of living in a world made for seeing people. Creating a visual assistance application would be a great way to improve the lives of many people from different ranges of blindness. By applying machine learning and existing technologies, it is possible to make use to already available devices to improve daily life. Currently, some examples of existing assistive technologies include portable magnifiers, braille, auditory or dictation related technology, synthetic speech, and more. Some of the most common devices used in life are phones and computers, which are already capable of implementing new advancements. This project takes advantage of their capabilities to create a new way to navigate in their daily lives.

Visual Impairment Background

Visual Impairment is a disability that can range anywhere from blurry vision to complete blindness. It is important to note that many visually impaired individuals may still have some form of vision even if it is limited. Accommodating for these individuals includes different techniques such as light and dark color contrasts in public spaces, implementing braille, and taking advantage of computer programs to assist in daily tasks.





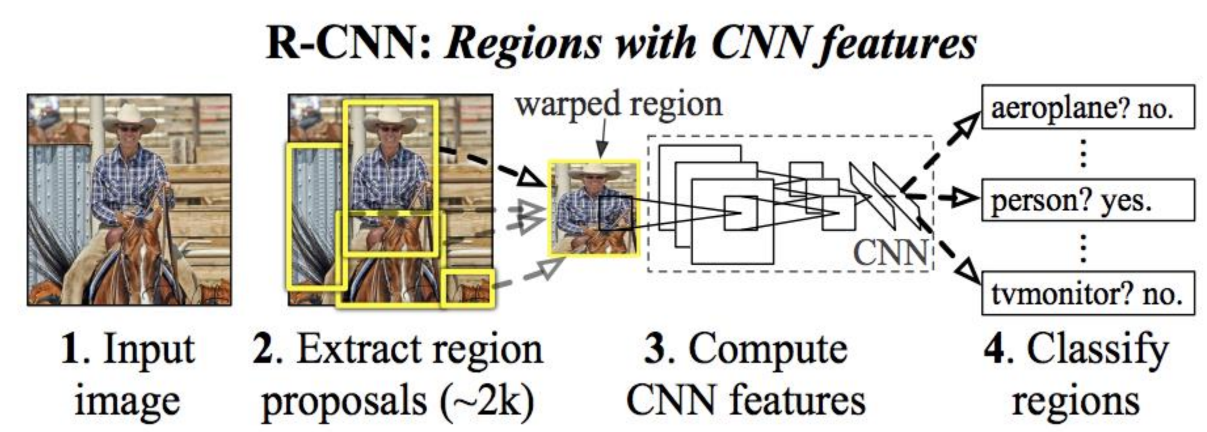


Key Terms

Visual Impairment - A term that encompasses both those who are blind and those with low vision.
Functional Limitation - The interaction of visual functioning and ability to perform activities of daily living/instrumental activities of daily living.
Low vision – A person who has measurable vision but has difficulty accomplishing visual tasks even with prescribed corrective lenses.
Vision Loss - Individuals who have trouble seeing, even when wearing glasses or contact lenses, as well as individuals who are blind or unable to see at all.
Legal Blindness – A level of vision loss that has been legally defined to determine eligibility for benefits.
Total Blindness - An inability to see anything with either eye.
Vision Loss - Individuals who have trouble seeing, even when wearing glasses or contact lenses, as well as individuals who are blind or unable to see at all.

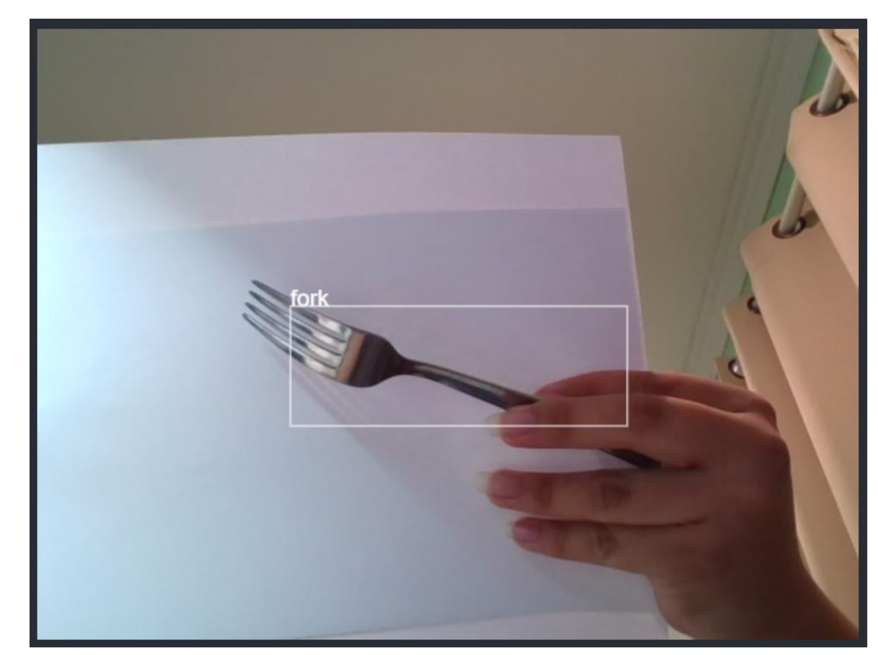
Computer Vision Methodology and Implementation

Computer vision uses object recognition, a subsection of machine learning. This machine learning uses classification to make predictions on what objects are seen on the screen. Through supervised learning and labeled datasets, the machine can learn and infer what an object can be based on the datasets provided. The specific subset of TensorFlow uses region-based convolutional neural networks (R-CNN). This is a deep learning approach that attempts to identify objects and their location on the screen. Please see the diagram below for a visual representation. Usually, objects that are recognized are contained within a bounded box where features are extracted to specifically highlight only one object at a time. The CNN that's used to compute the features specifically creates segmented masks that comprise of the main components of the current image. It would then use a Support Vector Machine to see if each object contained within a mask contains an instance of a trained object. As it currently stands, object recognition models cannot have 100% accuracy, so it is common for misreads on detected objects. Continuous improvements in model training and development can bring the accuracy much higher to become more reliable for applications such as this project.

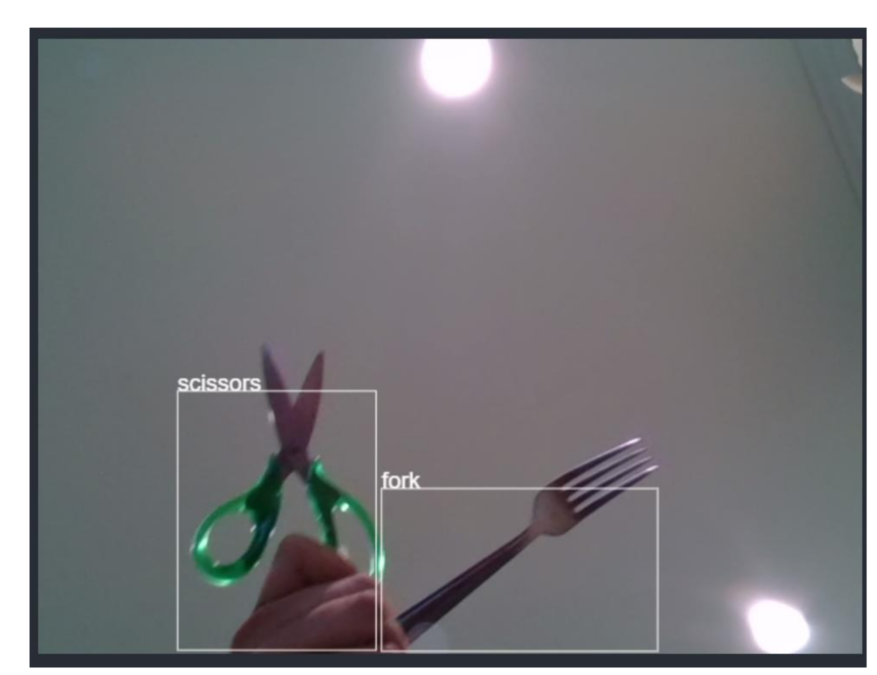


Results

The resulting project is a web application coded in React Native using the TensorFlow pre-trained model. It can recognize objects within the pre-trained library of TensorFlow but its current audio feedback capabilities are limited. All audio recordings are personally done by me and are added as assets for the project. To the right is a demonstration of what the screen will look like when an object is detected. The object is bounded within a box with the correct name of the object associated with it. Within the actual demo, only the matching audio will be played out to the user to let them know what objects are currently in front of the camera device. There are currently a limited number of recorded audio for feedback as I wanted to limited the sample size as well as ensure that the objects recorded are specifically ones that would be used in daily life. One aspect included for consideration of low vision individuals is limited usage of white text and boxes in the case that they can vaguely detect light and dark contrast.



As the project currently stands, it may have difficulties when presented with multiple objects on the screen. While it does detect all objects and plays the correct audio files, the timing of the objects as well as directionality regarding the placement of the object within the screen is limited. Within the provided demo video, there is a demonstration of detecting multiple objects where, although there is a delay for clearer feedback, the timing has not been properly calibrated to match large groups of detections. Certain improvements are still needed but with the two rounds of testing conducted. Participants were able to detect relevant objects with decent ease. There were reported inaccuracies which are more likely due to the training model and its predictions, but the audio provided mostly matched when it was necessary to play. To see a better video demonstration, please scan the QR Code below to view at your own leisure.
The Github with the code will also be provided for better understanding on the format and organization of the code and assets.



Demo Video | **Github Code**

Summary and Conclusions

This project was created to understand and find a possible solution to make the world more accessible for visually impaired individuals. Through computer vision, this application can detect objects and relay feedback to the user. Currently, the application can only relay audio feedback for a limited number of items but requires extra work to integrate as a mobile application. The working features include a webcam connection, real-time object detection, audio feedback for 13 items, and limited low-vision light-dark visual contrast markers. Training and development of the model and application is still possible to become more inclusive of day-to-day items. The status of the project is a proof of concept which still has possible improvements. From the testing and feedback, the main concept can be a useful tool but implementing it as a mobile application would make it more portable. It is important to consider other ways to access the application such as creating audio queues or including ways to repeat feedback to aid different needs for differing levels of assistance required.
Object recognition is an important technology that could aid in disability assistance. Whether through self-driving cars, security systems, or just applications within your phone, there are countless ways to apply this technology. This project helped point out important issues that visual impairment creates for people in their daily lives. Becoming independent and having resources available can make a more inclusive and safer environment for all. Developing these technologies are integral to the future of society. By investing into this growing field, we can work to live with technology and not against it.

Future Work

Future implementations of this project would include the integration of this technology into a fully-fledged mobile application. As this project uses React Native, translating the current framework into one suited for mobile applications is possible. In addition, considering other languages and devices would make this application more portable. Whether through other pretrained models like Yolov8, other new and faster implementations can be made. Possible improvements include making the low-vision visual contrast markers more apparent by highlighting the whole box, adding directionality to audio queues, and adjusting the brightness of objects on the screen. Going through extensive testing would enhance the user interface and include new features like natural language processing for verbal interactions. Enhancing usability would also include continuing to train the model to include a wider variety of home items.

References

"Assistive Technology for Blind or Low Vision Participants." *MIUSA*, 13 Mar. 2023, www.miusa.org/resource/tip-sheets/assistivetechnologyforblind/.

"How TensorFlow Object Detection Works and Where to Use It." *TF Certification*, TensorFlow, 22 Sept. 2021, www.tfcertification.com/blog/how-tensorflow-object-detection-works-where-to-use-it.

Kleck, Hillary. "What Blindness Really Looks Like." *Perkins School for the Blind*, 16 Oct. 2023, www.perkins.org/what-blindness-really-looks-like/.

Potrimba, Petru. "What Is R-CNN?" *Roboflow Blog*, Roboflow, 25 Sept. 2023, blog.roboflow.com/what-is-r-cnn/.

Acknowledgements

I would like to thank Dr. David Johannsen who offered insight into the many possibilities to conduct this project. His valuable advice and perspective helped me create a version of my project that showcased truly how helpful technology can be within the modern world. I would also like to thank my friends and family for supporting and aiding in the testing process. I would not have been able to complete this project without your help.