Mobile Application for Object Detection and Audio Feedback to Aid Visually Impaired Navigation

ABSTRACT

Visually impaired individuals often encounter challenges in navigating their surroundings due to the lack of visual information. This project presents a mobile application designed to enhance their mobility and independence by providing real-time object detection and audio feedback. Leveraging the YOLO (You Only Look Once) algorithm, a state-of-the-art object detection model, the application utilizes a mobile device's camera to identify objects in the user's environment. The system delivers precise audio feedback, such as "person in front of you" or "table to your right," describing the detected objects and their relative positions using text-to-speech technology. This hands-free solution offers an accessible and practical tool for visually impaired individuals, enabling safer and more confident navigation in various environments. The integration of object detection and auditory cues provides a robust assistive technology for addressing everyday challenges.

Keywords:
Mobile Application, Object Detection, YOLO, Visually Impaired, Audio Feedback, Text-to-Speech.

Chapter 1

Introduction

Navigating everyday environments poses significant challenges for visually impaired individuals, as they rely primarily on auditory and tactile senses to understand their surroundings. Traditional mobility aids, such as white canes and guide dogs, offer valuable assistance but come with limitations. White canes, for instance, can detect obstacles only within close proximity, while guide dogs require extensive training and may not be accessible to everyone. These limitations highlight the need for innovative technological solutions that can enhance spatial awareness and provide real-time information about the surrounding environment.

With the rapid advancements in artificial intelligence (AI) and computer vision, mobile applications have emerged as powerful assistive tools. This project introduces a mobile application designed to aid visually impaired individuals by integrating real-time object detection with auditory feedback. The application utilizes the YOLO (You Only Look Once) algorithm, a state-of-the-art deep learning model known for its speed and accuracy in detecting objects. By leveraging a smartphone's camera, the system continuously scans the environment and identifies objects within the user's vicinity. Once an object is detected, the application provides immediate audio feedback using text-to-speech technology, offering intuitive and descriptive cues such as "person ahead," "chair to your left," or "door in front of you." This enables users to gain a better understanding of their surroundings without requiring physical interaction.

Unlike conventional assistive devices, this mobile-based solution is lightweight, cost-effective, and accessible to a wider audience, requiring only a smartphone with a camera and internet connectivity for initial model deployment. By delivering precise and real-time auditory guidance, the application enhances the mobility and independence of visually impaired individuals, enabling them to navigate unfamiliar spaces with greater confidence and safety. The integration of AI-powered object detection and speech synthesis technology transforms the way visually impaired individuals interact with their surroundings, offering a seamless and intuitive assistive solution tailored for real-world applications

Chapter 2: Review of Past Work and Problem Formulation

2.1 Review of Past Work

Object detection has been a prominent area of research in computer vision, contributing significantly to fields such as autonomous driving, security, healthcare, and industrial applications. Over the years, various algorithms have been proposed to address the challenges inherent in detecting and classifying objects within images. These algorithms vary in terms of accuracy, speed, computational complexity, and ease of implementation, making the selection of an appropriate model for real-time applications a complex task.

One of the most influential advancements in object detection is the emergence of deep learning-based techniques, which outperform traditional methods that relied on manual feature extraction. Early algorithms such as Haar cascades and Histogram of Oriented Gradients (HOG) were designed to detect objects through feature extraction followed by classification. While these methods were effective in controlled environments, they had significant limitations in terms of accuracy, scalability, and robustness when dealing with noisy data or varying conditions.

With the advent of Convolutional Neural Networks (CNNs), a paradigm shift occurred in object detection. The breakthrough with AlexNet in 2012 demonstrated the power of CNNs in visual recognition tasks, which spurred the development of several CNN-based object detection models. These include Single Shot Multibox Detector (SSD), Faster R-CNN, and You Only Look Once (YOLO).

1. **YOLO (You Only Look Once)**: YOLO represents a single-stage object detection model that is designed for real-time applications. Unlike traditional models that process images in multiple stages, YOLO performs detection in a single pass. By dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell, YOLO can quickly identify objects in real-time. However, while YOLO offers high speed, it sacrifices some accuracy when compared to more complex models like Faster R-CNN. Despite this tradeoff, YOLO remains a leading choice for real-time applications due to its efficient processing speed.

2. **SSD (Single Shot Multibox Detector)**: SSD strikes a balance between speed and accuracy. It also operates as a single-stage detector but improves upon YOLO by predicting detections at multiple feature scales, allowing it to detect smaller objects more effectively. SSD performs well in scenarios where both speed and accuracy are important. However, like YOLO, SSD may not achieve the same level of precision as two-stage detectors.

3. **Faster R-CNN**: Faster R-CNN is a two-stage detection model that emphasizes accuracy over speed. The first stage involves generating region proposals through a Region Proposal Network (RPN), while the second stage classifies these proposals into object categories and refines the bounding boxes. Faster R-CNN consistently delivers high accuracy and is particularly effective in environments with complex backgrounds or where object details are critical. However, its relatively slow processing speed limits its applicability in real-time applications, making it less suitable for tasks such as autonomous driving or live surveillance.

In summary, the comparative analysis of these object detection algorithms reveals a fundamental trade-off between speed and accuracy. While YOLO excels in real-time performance, SSD offers a more balanced solution, and Faster R-CNN delivers superior accuracy at the expense of speed. These insights are crucial for practitioners selecting a suitable model for their specific use cases, especially when considering the deployment of object detection systems in real-world environments with varying constraints.

2.2 Problem Formulation

Despite the significant advancements in object detection, several challenges remain in applying these algorithms to real-world applications. The core problem lies in selecting an optimal algorithm that meets the specific requirements of a given use case, particularly when there is a trade-off between speed and accuracy.

The primary goal of this research is to explore the capabilities and limitations of the YOLO algorithm for object detection in practical applications, such as autonomous driving, security surveillance, and industrial inspection. The focus is on determining whether YOLO, as a real-time detector, can be effectively used in scenarios that require high speed without compromising too much on accuracy.

By focusing on YOLO's strengths and addressing its weaknesses, this study aims to contribute to the growing body of knowledge on real-time object detection algorithms and offer practical insights for the development of efficient object detection systems.