International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Advanced Edge AI Navigation and Assistance System for Blind User

Divya[1], Rajapandi[2]

[1,2]Dept of Computer Science and Business Systems & Sethu Institute of Technology, Pulloor, Kariapatti – Virudhunagar 626 115

Abstract

Despite substantial advancements in visual assistance technology, many existing systems are limited by sensor capabilities, computational resources, and power consumption. Traditional computer vision algorithms struggle to perform complex tasks required for real-time navigation. This paper presents the design and implementation of an advanced computer vision-based navigation system specifically tailored for Blender users, aimed at facilitating independent navigation in diverse environments .By leveraging edge Artificial Intelligence (AI) and deep learning methodologies, the proposed system achieves real-time object detection, person recognition, and environmental awareness, addressing the critical challenges posed by conventional systems. The use of cost-effective, low-power mobile computing platforms, such as smart depth sensors like the OpenCV AI Kit-Depth (OAK-D), enables efficient processing while ensuring portability. This system not only enhances the ability to identify and navigate obstacles but also incorporates additional functionalities, including reading written notices aloud and responding to traffic signals. Key design considerations, such as training data collection, computational efficiency, and portability, have been meticulously addressed to ensure reliable performance in real-world scenarios. The incorporation of an AI-driven voice interface facilitates user-friendly interaction, making this system an innovative and unobtrusive visual assistance device.

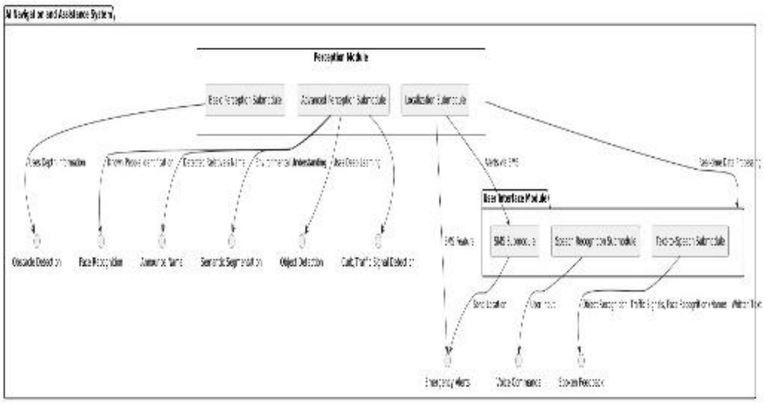
Keywords: Edge Artificial Intelligence (AI), Mobile Computing, Visual Assistance Device, Deep Learning.

1. INTRODUCTION

Visual impairment remains a pressing global issue, affecting millions of individuals worldwide. According to the World Health Organization, approximately 285 million people were visually impaired in 2010, with 246 million experiencing low vision and 39 million classified as completely blind. The demographic most affected comprises older adults, with 65% of the visually impaired and 82% of the blind being 50 years of age or older. Major contributors to visual impairment include uncorrected refractive errors (43%) and cataracts (33%). The causes of blindness are similarly concerning, with cataracts accounting for 51% and other conditions such as glaucoma and age-related macular degeneration contributing to the statistics.Despite significant progress in reducing the prevalence of visual impairment—from 4.58% in 1990 to 3.38% in 2015—challenges remain. An aging population continues to drive an increase in the total number of visually impaired individuals, with projections suggesting that by 2030, the number will rise to approximately 385 million globally, including 55 million who will be legally blind.Visually impaired individuals face numerous challenges, including reliance on others, unemployment, reduced social interaction, difficulties with reading and writing, and barriers in transportation and device usage. Current solutions such as guide dogs, walking canes, and GPS-enabled devices each have limitations, particularly in outdoor navigation where accurate assessment of traffic and road conditions is critical. While guide dogs can identify obstacles, their communication with human users can be ambiguous, and walking canes excel at detecting ground-level hazards but often miss overhanging obstacles. GPS devices assist with routing but fall short in obstacle detection.In response to these challenges, we propose an advanced computer vision-based navigation system designed specifically for Blender users, aimed at enhancing their independence and mobility. Our system employs edge AI technology and deep learning to provide real-time object detection, person recognition, and environmental awareness. By integrating functionality to read written notices aloud and respond to traffic signals, this innovative system addresses the multifaceted needs of visually impaired individuals, enabling them to navigate their surroundings safely and confidently.The remainder of this paper is structured as follows: Section 2 reviews related work and existing technologies; Section 3 details the design and functionality of our proposed navigation system; Section 4 describes the hardware and setup; Section 5 presents performance evaluation results; and Section 6 concludes with future enhancements and potential extensions of this project

2. Related Work

Visual assistance devices are commonly classified as Electronic Travel Aids (ETAs) that provide information about the environment through a convenient user interface. Various approaches have been proposed in the literature for the design of visual assistance systems based on the underlying sensory systems, hardware configuration, physical setup, data inference techniques and user interface. The most used sensor types include ultrasound, sonar, laser, RGB CCD camera, infrared (IR) camera and GPS. Some approaches convert the input sensor data to other modalities [11], [12]. For the user interface, audio transmitted via earphones or hand gloves equipped with buzzers or tiny vibrating actuators are typically used. Early visual assistance system designs were based on projecting a camera image onto the human skin using vibrating motors [10] and sensor modality conversion, where ultrasonic waves were converted to the audible range and the converted audio was used to understand the environment [11], [12]. Although the vOIce system [12], where visual image data are converted to human audible frequency, showed promising results these systems were typically slow, physically uncomfortable, and obtrusive, provided only very coarse information about the surroundings and required extensive user training to be used effectively. Early visual assistance systems based on GPS data [13], focused primarily on navigation (i.e., neither collision avoidance nor obstacle detection was performed) and often suffered from signal loss, especially in indoor environments and urban areas. Visual assistance systems based on RFID technology provided good localization in indoor environments wherein RFID tags were physically placed [14]. Since RFID sensing methods provide a range rather than an accurate geolocation of the tags, the resulting localization errors were unacceptable in certain situations. In the past few decades, as the sensor and computing technologies have evolved remarkably, so have visual assistance systems. NavBelt [15] is a real-time visual assistance system that uses eight ultrasonic sensors mounted on a waist belt worn by the user with a computing unit located in a backpack. GuideCane [16] is a wheel based cane with an ultrasonic sensor attached to a main processing unit that is mounted on the cane. Bousbia-Salah et al. [17] describe a visual assistance system that uses two ultrasonic sensors strapped onto the user’s shoulders accompanied by an accelerometer, with a foot switch for error control. The CyARM system [18] consists of a handheld device with two ultrasonic sensors to detect obstacles, coupled with a wire-enabled user interface mounted through a waist belt. While ultrasonic sensors are low-cost and fast, they fail to provide accurate geometric descriptions of the obstacles encountered and are prone to errors caused by noise and signal reflections. Extensive reviews of ETAs and the challenges faced in their design and deployment can be found in [19] and [20]. Recently, promising advancements in the design of ETAs have been made using computer vision-based approaches. Tapu et al. [21] use a smartphone camera mounted on a chest harness to detect moving obstacles using the multiscale Lucas-Kanade feature tracking algorithm. Object detection is performed by classifying a Bag of Visual Words (BoVW) and Histogram of Oriented Gradients (HOG) features using linear classifiers such as the Support Vector Machine (SVM) and image ranking methods. Jabnoun et al. [22] propose an ETA that uses SIFT feature-based object detection on video streams. These ETAs employ traditional computer vision methods for object detection that are not robust in real-world environments. Moreover, they lack 3D information, in that the user does not know how far away the obstacles are. The Electron-Neural Vision System (ENVS) uses eyewear-based stereo cameras to obtain 3D descriptions of the environment from depth images [23]. The system is also equipped with a GPS, portable computer, Transcranial Electrical Nervous Stimulation (TENS) unit and a TENS based glove as the user interface. Rodriguez et al. [24] propose a stereo vision-based system where plane segmentation is performed to detect ground pixels and a polar grid notation used to detect obstacles within depth images. The stereo cameras are mounted in the chest region, coupled with an audio-based user interface. Johnson and Higgins [25] proposed a stereo vision-based scheme, with stereo cameras mounted in the hip region on a tactile belt with 14 vibrating motors worn by the user, that also serves as a user interface. While the above systems provide advanced, accurate 3D perception of the environment and have robust, reliable obstacle detection capabilities by exploiting stereoscopic vision, they lack the scene understanding capabilities needed for assessment of traffic, road and sidewalk conditions, and advanced deep learning based perception capabilities, such as image classification, objection detection and semantic image segmentation. There has been significant recent progress in the design and development of visual assistance systems that derive a richer understanding of the user's environment. State-of the-art visual assistance systems are designed to use a variety of sensor types including conventional CCD RGB cameras, stereoscopic cameras, GPS-enabled devices, and ultrasound and RFID sensors. Several systems incorporate sensor modality conversion techniques in conjunction with an intuitive user-friendly interface. Existing systems that use camera-based sensors typically employ traditional computer vision algorithms on predefined image features such as SIFT, HOG and BoVW, to mention a few, that are observed to perform adequately in constrained, well defined environments but fail to generalize to a real-world settings characterized by greater variability. Also, existing camera-based systems do not adequately tackle advanced perception tasks such as assessment of traffic conditions and elevation changes (e.g., curb detection), understanding of traffic signs, detection of crosswalks, etc. Several conventional camera-based systems have a bulky, obtrusive physical setup, which is not user-friendly and may draw unwarranted attention in public spaces. Consequently, there is a need for visual assistance systems that execute deep learning-based inference algorithms for advanced perception tasks, while preserving a simple, unobtrusive form factor and consuming sufficiently low battery power. This would ensure user mobility and avoid unwarranted attention during user ambulation in public spaces. In this paper we propose a novel and practical design for a visual assistance system using a smart stereo vision sensor called OAK-D and edge AI devices that can overcome the limitations of existing approaches by using edge AI technology. We propose a visual assistance system with deep learning capabilities to perform advanced computer vision tasks, such as object detection and semantic image segmentation in real time using edge AI devices such as the Intel's NCS2. The proposed system is designed to perform sophisticated scene understanding tasks such as detection of roads, curb entry/exit locations, crosswalks, assessment of traffic conditions by detecting traffic lights and congestion, reading of traffic signs and street names using state-of-the art computer vision and deep learning methods. The proposed system can perform 3D point cloud processing to detect elevation changes at curb entry/exit locations. The proposed system is equipped with a GPS-enabled device for geolocalization that can save custom locations for convenience. The GPS coordinates along with a snapshot of the current location can be shared with preferred contacts over a short message service (SMS) if the user needs emergency assistance. A user-friendly, customizable, AI-based interactive voice interface equipped with speech recognition provides the user periodic updates about the environment. The physical setup consists of a chest-mounted OAK-D sensor placed inside a vest and connected to a computing unit placed in a small backpack. Wireless Bluetooth earphones with a microphone are used for voice-based interactions. The setup is unobtrusive and not noticeable as an assistive system. Interviews were conducted with visually impaired people, including those from non-profit organizations such as LightHouse for the Blind Inc., to comprehend and catalog the common daily challenges they face. We prioritized and ranked the identified challenges based on these interviews and addressed them in our design. i added related work for my project change the words and prasers this related work short this and give effectrive



In the **Primitive Perception** submodule, obstacles are detected within a certain range to alert the user and prevent collisions. The focus is on fast and efficient detection, relying on computer vision techniques rather than depth sensors like OAK-D. A standard camera feed is divided into multiple regions of interest (ROIs), each representing different areas around the user (left, right, center, top). These regions are used to detect obstacles, including hanging objects like wires or overhead barriers.Obstacle detection is done by analyzing pixel movement and changes in frame sequences. Algorithms like optical flow or contour detection are used to identify potential hazards. Once detected, each obstacle's position and proximity are calculated using geometric approximations. The system marks the grid cells where obstacles are detected, ensuring comprehensive coverage around the user.The **Primitive Perception** submodule offers real-time obstacle detection across various shapes and sizes, providing timely alerts to the user. The system ensures a high inference rate, leveraging the host machine's CPU and OpenCV libraries for efficient processing. Sample outputs include bounding boxes around detected obstacles, displayed alongside the real-time camera feed foreasy visualization.

3.2. Advanced Perception

The **Advanced Perception** submodule performs comprehensive scene understanding tasks, such as image classification, object detection, and semantic segmentation, to extract detailed perceptual information about the environment. Unlike simpler systems, this submodule aims to provide a deeper and more nuanced understanding of the user's surroundings, identifying objects, reading signs, and recognizing visual cues that help with navigation.This system is designed with constraints similar to those faced by developers of autonomous systems, including limits on computational complexity, battery consumption, and hardware cost. We leverage advanced computer vision and deep learning techniques optimized for edge AI platforms, avoiding the need for expensive GPU-based solutions. Model optimization techniques, such as OpenVINO and TensorFlow Lite, are applied to run deep learning models on low-cost, small-form-factor devices with real-time inference capabilities.

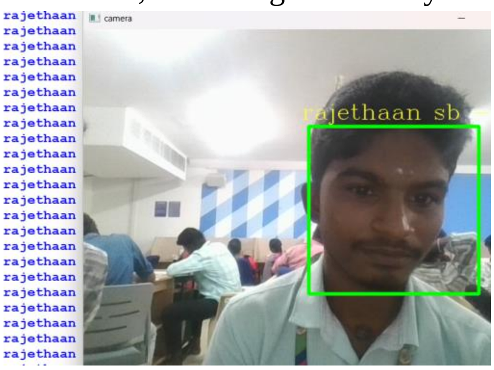
Key tasks handled by the **Advanced Perception** submodule include:

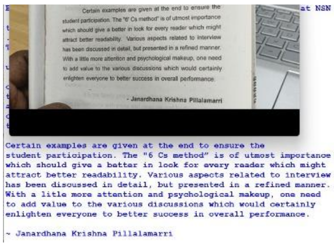
- **Object Detection**: Using pre-trained models and custom-trained deep learning models, the system can detect and classify objects such as people, vehicles, bags, books, traffic signs, and other items relevant to the user's environment. This functionality is crucial for identifying obstacles, reading visual cues, and ensuring a safe path for the user.
- **Text Detection and Reading**: The system is capable of detecting written text from signs, books, or notices and converting it into speech. This feature helps users understand and respond to important visual information in their surroundings, such as street names, traffic signs, or notices.
- **Traffic Signal and Curb Detection**: Deep learning models trained to recognize traffic signals, curbs, and road markings are deployed to help users navigate safely. The system alerts the user to changes in traffic light colors and indicates safe or hazardous paths.
- **Facial Recognition**: The system can detect and recognize familiar faces (e.g., family members or acquaintances) and provide the user with an audio notification when a known person is approaching.

For model training, publicly available datasets such as the **COCO** dataset and **Open Images Dataset V6** were used, along with custom data collected for specific tasks like object detection and text recognition. The models have been fine-tuned with images and video data from real-world environments to optimize accuracy and performance in varied lighting and weather conditions.

The deep learning models are optimized for execution on mobile and edge computing devices using model compression techniques to ensure efficient use of resources while maintaining accuracy. For instance, models are quantized to reduce their memory footprint and improve inference speed without significantly sacrificing performance. By employing these techniques, the system can achieve real-time performance with a low-power standard camera and modest processing hardware.

This approach allows the **Advanced Perception** submodule to deliver robust scene understanding, facilitating tasks such as identifying obstacles, reading traffic signs, and recognizing people in real time. The system integrates seamlessly with the rest of the AI Navigation and Assistance platform to offer comprehensive and timely feedback to users, enhancing their ability to navigate independently and safely.





Here's an object detection description tailored to your project, structured similarly to the reference:

1. Object Detection

For the object detection component of the project, we implemented a customized deep learning pipeline capable of detecting relevant objects such as books, bags, mobile phones, mice, fans, and other personal objects. This system is essential for providing visually impaired users with real-time awareness of the objects in their environment, enabling safe and independent navigation.

We utilized the SSD-MobileNet object detection model, pre-trained on the COCO dataset, for its balance of speed and efficiency. The SSD-MobileNet model is particularly suited to low-power edge devices due to its compact architecture and ability to perform single-stage detection. Custom models were also trained to identify specific classes relevant to the user's surroundings, such as personal objects (e.g., phones, books) and general objects (e.g., fans, mice, bags).

Model Training and Setup

The model was trained on various image resolutions ranging from 640×480 pixels to 300×300 pixels over 350,000 steps, using a batch size of 32. We adopted a training-to-validation-to-test data split of 70:20:10. The initial learning rate was set to 0.001 with a decay rate of 0.95. Our best results were achieved with a resolution of 300×300 pixels, yielding a mean average precision (MAP) of 0.68, which is suitable for real-time object detection in dynamic environments.

Despite its strengths, lightweight models such as SSD-MobileNet often suffer from a trade-off between speed and accuracy. To address the occurrence of false negatives (FN) (missed detections) and false positives (FP) (incorrect classifications), we lowered the model's confidence threshold to 0.35, which slightly increased the FP rate for non-target background objects, such as walls or shadows.

Handling False Positives and False Negatives

To mitigate the increase in false positives, we trained an additional lightweight CNN-based image classifier that was added to the pipeline. This classifier consists of:

A CONV=>RELU=>BN=>POOL layer, followed by

2 sets of (CONV=>RELU=>CONV=>RELU) layers with pooling layers between them.

This secondary classifier refines the object detection results by filtering out false positives, such as distant background objects. Further improvement was made by introducing a temporal filter: any detection that persisted for less than 2 seconds was considered a weak detection and discarded. This additional step

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

helped suppress transient false positives and improved the overall accuracy of the system. The object detection pipeline is structured as shown in (below), detailing the flow from initial object detection to the final, filtered output.

Results and Performance

Our optimized pipeline effectively handles object detection tasks relevant to the user's environment, enabling real-time identification and feedback on objects such as phones**,** books**,** bags**,** and fans. The combination of SSD-MobileNet and the custom CNN classifier ensures that the system provides timely and accurate notifications about objects in the user's vicinity. This capability is crucial for enhancing independent navigation and daily interactions for visually impaired users.

3.2.2 Semantic Scene Understanding

While object detection is critical for identifying individual items, it alone is insufficient for full scene understanding, especially in environments where users need to distinguish between books, notices, or other written information, and navigate through indoor and outdoor spaces. To address this, semantic image segmentation models are deployed to understand the spatial layout and context of the scene.

In our system, semantic segmentation is used to provide an enhanced level of environmental understanding by identifying key areas, such as book pages**,** notices**,** signs**,** and other visually important cues. This ensures that visually impaired users receive a more comprehensive overview of their surroundings, facilitating navigation and the reading of written materials.

Model Selection and Optimization

For real-time performance on edge AI platforms like OAK-D and NCS2, we opted for lightweight models optimized for resource-constrained environments. Unlike heavyweight models like UNet **or** PSPNet**,** which require powerful GPUs for real-time inference, our project utilizes efficient models that can run on low-power devices.

We explored OpenVINO's pre-trained semantic segmentation models, such as semantic-segmentation-adas-0001 and road-segmentation-adas-0001**,** which are designed for autonomous driving tasks but repurposed for our system. These models were chosen for their balance of performance and efficiency.

The semantic-segmentation-adas-0001 model, though constrained in size (FP16), was ideal for outdoor scenarios, accurately identifying larger objects and signage. However, due to its complexity, this model could not run on the OAK-D or NCS2 and achieved a processing speed of approximately 2.3 fps on a standard CPU. For more resource-efficient segmentation, we deployed the road-segmentation-adas-0001 model, which can run on multi-Myriad platforms (OAK-D and NCS2). This model was used to segment key areas such as background**,** written material, and furniture in indoor settings, and achieved higher accuracy in recognizing real-time contextual elements.

For indoor semantic scene understanding, TensorFlow Lite's DeepLabv3 Mobile Net, pre-trained on the ADE20K dataset, yielded the best results, especially for detecting indoor environments such as walls**,** furniture, and written notices. However, this model was slower, running at about 0.4 fps due to its heavy computational requirements.

Model Challenges and Solutions

A significant challenge in using pre-trained models is that they are often trained on datasets collected under specific conditions, which might not fully generalize to all environments. For instance, some models struggled to differentiate between objects in varying indoor lighting conditions or missed detecting written text in complex layouts. To mitigate this, additional fine-tuning was performed on a custom dataset consisting of books**,** notices, and objects in diverse lighting and angles, improving overall accuracy in real-world scenarios.

The system also incorporates a confidence filter, ignoring weak detections and providing feedback only on clearly identified areas, ensuring a high level of accuracy when guiding the user through their environment.

Results and Performance

The semantic scene understanding submodule is vital for providing visually impaired users with a detailed comprehension of their surroundings. Whether it's helping to read signs, detect objects in complex environments, or differentiate between written content and background elements, the system ensures a reliable understanding of the environment in both indoor and outdoor settings. The seamless integration of object detection and segmentation models significantly enhances the user's navigation experience, ensuring safety and independence in varied environments

3.2.3 Crosswalk detection

In our project, crosswalk detection is achieved using advanced computer vision techniques tailored for dynamic outdoor environments. Traditional methods like edge detection and Hough transform, often used in road lane detection, are not suitable due to the varying camera angles and the complex, noisy textures of the road surfaces encountered during use. Instead, we employ semantic segmentation models optimized for edge AI platforms, such as the road-segmentation-adas-0001 model. This model provides more reliable results under diverse lighting conditions and road textures compared to conventional image processing routines.

To enhance accuracy, the model's predictions undergo further processing, including noise removal and contour analysis. By analyzing the area, convex hull, orientation, and solidity of each connected component, the system can accurately identify crosswalks even in challenging scenarios, such as faded lines or obstructed road markings. The chosen contours are those that align with specific orientation angles and solidity values, ensuring that the detected crosswalk is relevant to the user's current viewpoint.

This approach is computationally efficient, as crosswalk detection is activated only within a certain spatiotemporal range after the system identifies stop signs, reducing unnecessary processing. The effectiveness of this method has been demonstrated in various outdoor settings, with the chest-mounted camera configuration yielding the best results due to the clearer capture of road textures. This advanced detection mechanism plays a crucial role in enabling safe and independent navigation for users in real-world environments.

Here is the revised version:

4. System Hardware Description

The proposed system is designed to be simple, unobtrusive, and easily wearable, eliminating the need for handheld devices such as a cane or laser sensor. The setup includes a small backpack that houses a compact host computing unit, such as a Raspberry Pi, Chromebook, or laptop, along with a battery power unit. An OAK-D sensor is mounted in the chest area inside a vest. The sensor's position can vary without the need for calibration, offering flexibility in its placement. The sensor is then connected to the computing unit in the backpack.

The OAK-D sensor is selected for its built-in AI processing capabilities, allowing most computer vision tasks to be processed directly on the sensor chip. This data is then transmitted to the host computing unit, which provides real-time updates to the user through a voice interface via Bluetooth-enabled wireless earphones. A USB-enabled GPS device is attached to the host computing unit and is mounted on the exterior of the backpack. The OAK-D sensor is powered by a compact 10,000 mAh battery pack, offering up to 8 hours of operation. A five-year-old Lenovo Yoga laptop with 8GB RAM and an Intel i7 processor is used as the host computing unit, supplemented by a neural compute stick.

5. Experimental Evaluation and Results

We targeted lightweight models such as the SSD MobileNet for object detection to ensure higher inference rates of approximately 30 frames per second (fps) using the OAK-D sensor. This configuration enabled us to run multiple object detection models simultaneously on the CPU, NCS2, and OAK-D, allowing real-time detection of as many object classes as possible.

The traffic signs were classified using the TrafficSign Net model, which demonstrated promising performance metrics. The model was trained with a decaying learning rate of 0.0001, a batch size of 64, and for a total of 100 epochs. Popular traffic datasets such as LISA and GTSRB were utilized alongside our custom dataset. The detections were subsequently mapped to the depth image to compute their distances from the user.

After converting the Traffic Sign Net model to the Open VINO format, we achieved inference rates exceeding 60 fps on the OAK-D sensor and around 100 fps on the CPU. The SSD Mobile Net model maintained an inference rate of approximately 30 fps. Given that the OAK-D does not support multi-object detection models, the traffic object detection tasks were executed on the external NCS2 devices.

Overall, our system, which integrates primitive perception, object detection models, and classification algorithms, shows satisfactory performance for obstacle detection. However, running semantic image segmentation models simultaneously with other models presents challenges. None of the semantic image segmentation models were able to operate on a single NCS2 device. The road-segmentation-adas-0001 model is the fastest among them but requires a CPU or multiple NCS2 devices to function effectively.

When combined with the entire system, which includes this model along with the primitive perception and object detection pipeline, we achieved an inference rate of around 10 fps. In contrast, other semantic image segmentation models displayed high computational complexity, leading to performance bottlenecks. For instance, models such as semantic-segmentation-adas-0001 and DeepLabV3 MobileNet achieved inference rates of 1.56 fps and 0.42 fps, respectively, while necessitating a host CPU for operation. Consequently, the overall system performance suffered when these heavyweight models were deployed alongside the other models. At present, these heavyweight models are used only upon user request, and the user is required to remain stationary for safety during their operation.

6. Conclusions and Future Work

In this project, we developed a novel and comprehensive vision system for the visually impaired, facilita-

tating both indoor and outdoor navigation while incorporating scene understanding. The system is designed to be simple, fashionable, unobtrusive, and discreet, making it less noticeable as an assistive device. It effectively
addresses common challenges such as detecting traffic signs, identifying hanging obstacles, recognizing crosswalks, navigating moving obstacles, detecting elevation changes, and geolocation through advanced perception capabilities implemented on a low-power device. The user-friendly voice interface enables seamless control and interaction with the system.
After conducting several hours of testing in Monrovia, CA, we are confident that this project effectively tackles the most prevalent challenges faced by visually impaired individuals in their daily navigation.
Looking ahead, our future work will focus on several enhancements, including GPS integration and AI interaction. With GPS capabilities, the system will provide precise, location-based instructions to users, significantly improving their navigation experience. Additionally, we aim to incorporate an interactive AI feature that allows users to ask questions about their surroundings. For example, while on the road, a user could inquire, "What should I do while on the road?" The system will respond with relevant guidance, ensuring a safer and more informed navigation experience.
We will also explore options to run multiple semantic image segmentation models simultaneously at higher inference rates. At the time of this study, the OAK-D sensor was a Kickstarter project, which limited our ability to acquire multiple sensors for testing. We plan to evaluate the system's performance with multiple OAK-D sensors operating concurrently. Additionally, we will experiment with traditional point cloud methods to detect elevation changes using the Generation 2 DepthAI module and aim to incorporate robust object tracking across frames for more accurate traffic analysis.
More importantly, based on insights gained from previous studies, we are confident that in our future iterations, we can eliminate the need for a laptop altogether. Instead, we envision utilizing mobile devices such as the Google Pixel 2 in conjunction with low-power computing edge devices like the Nvidia Jetson or TX2, significantly enhancing the system's mobility and user accessibility

References:

1. Ackland, P., Resnikoff, S., & Bourne, R. (2017). World blindness and visual impairment: Despite many successes, the problem is growing. Community Eye Health, 30(100), Article 100.
2. AI Benchmark. Detailed ranking processors. Retrieved April 18, 2021, from https://ai-benchmark.com/ranking_processors_detailed.html
3. Bourne, R. R., et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. Lancet Global Health, 5(9), Article 9.
4. Bourne, R. R., et al. (2020). Global prevalence of blindness and distance and near vision impairment in 2020: Progress towards the Vision 2020 targets and what the future holds. Investigative Ophthalmology & Visual Science, 61(7), Article 7.
5. Bousbia-Salah, M., Redjati, A., Fezari, M., & Bettayeb, M. (2007). An ultrasonic navigation system for blind people, pp. 1003–1006.
6. Collins, C. C. (1970). Tactile television-mechanical and electrical image projection. IEEE Transactions on Man-Machine Systems, 11(1), Article 1.
7. Dakopoulos, D., & Bourbakis, N. G. (2009). Wearable obstacle avoidance electronic travel aids for blind: A survey. IEEE Transactions on Systems, Man, and Cybernetics: Part C (Applications and

Reviews), 40(1), Article 1.
8. DepthAI: Embedded Machine learning and Computer vision API. Luxonis.
9. Daily Life Problems Faced by Blind People. (2021, February 25). Retrieved from https://wecapable.com/problems-faced-by-blind-people/
10. Ignatov, A., et al. (2019). AI benchmark: All about deep learning on smartphones in 2019. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (pp. 3617-3635).
11. Ignatov, A., et al. (2019). AI benchmark: Running deep neural networks on android smartphones. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (pp. 0-0).
12. Ito, K., et al. (2005). CyARM: An alternative aid device for blind persons, pp. 1483–1488.
13. Jabnoun, H., Benzarti, F., & Amiri, H. (2014). Visual substitution system for blind people based on SIFT description, pp. 300–305.
14. Johnson, L. A., & Higgins, C. M. (2006). A navigation aid for the blind using tactile-visual sensory substitution, pp. 6289–6292.
15. Kay, L. (1964). An ultrasonic sensing probe as a mobility aid for the blind. Ultrasonics, 2(2), Article 2.
16. Kulyukin, V., Gharpure, C., Nicholson, J., & Pavithran, S. (2004). RFID in robot-assisted indoor navigation for the visually impaired, pp. 1979–1984.
17. Loomis, J. M. (1985). Digital map and navigation system for the visually impaired. Unpublished manuscript, Department of Psychology, University of California, Santa Barbara.
18. Meijer, P. B. (1992). An experimental system for auditory image representations. IEEE Transactions on Biomedical Engineering, 39(2), Article 2.
19. Munemo, E., & Tom, T. (2013). Problems of unemployment faced by visually impaired people. Greener Journal of Social Sciences, 3(4), Article 4.
20. Real, S., & Araujo, A. (2019). Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. Sensors, 19(15), Article 15.
21. Rodríguez, A., et al. (2012). Assisting the visually impaired: Obstacle detection and warning system by acoustic feedback. Sensors, 12(12), Article 12.
22. Rosebrock, A. (2019). Traffic Sign Classification with Keras and Deep Learning. Retrieved from https://www.pyimagesearch.com/2019/11/04/traffic-sign-classification-with-keras-and-deep-learning/
23. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks, pp. 4510–4520.
24. Shoval, S., Borenstein, J., & Koren, Y. (1998). The Navbelt-A computerized travel aid for the blind based on mobile robotics technology. IEEE Transactions on Biomedical Engineering, 45(11), Article 11.
25. Tapu, R., Mocanu, B., & Zaharia, T. (2013). A computer vision system that ensures the autonomous navigation of blind people, pp. 1–4.
26. Ulrich, I., & Borenstein, J. (2001). The GuideCane-applying mobile robot technologies to assist the visually impaired. IEEE Transactions on Systems, Man, and Cybernetics-Part Systems and Humans, 31(2), Article 2.
27. Vu, H. T. V., Keeffe, J. E., McCarty, C. A., & Taylor, H. R. (2005). Impact of unilateral and bilateral vision loss on quality of life. British Journal of Ophthalmology, 89(3), Article 3.

28. WHO. (2021). Global Data on visual impairments 2010. Retrieved February 21, 2021, from https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf
29. Zhi-Han, L., Hui-Yin, Y., & Makmor-Bakry, M. (2017). Medication handling challenges among the visually impaired population. Archives of Pharmacy Practice, 8(1), Article 1.