RESEARCH & INNOVATION DIVISION
IJIST

Visually: Assisting the Visually Impaired People Through AI-Assisted Mobility

Muhammad Arsalan Kamran*, Alishba Orakzai, Umama Noor, Yasir Saleem Afridi, Madiha Sher.
Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar, Pakistan.
Correspondence: Muhammad Arsalan Kamran arsalankamran80@gmail.com

Citation | Kamran. M. A, Orakzai. A, Noor. U, Afridi. Y. S, Sher. M, "Visually: Assisting the Visually Impaired People Through AI-Assisted Mobility", IJIST, Special Issue pp 9-17, May 2024
Received | May 05, 2024, **Revised |** May 10, 2024, **Accepted |** May 16, 2024, **Published |** May 20, 2024.

This research introduces "Visually", a revolutionary mobile application that aims to address the complications that visually impaired people come across in their daily lives. By deploying advanced deep learning models for real-time object detection, facial recognition, and currency identification with voice outputs for each feature, the "Visually" application strives to enhance the autonomy, independence, and mobility of visually impaired people. The system undergoes thorough training on a diverse dataset, incorporating augmentation techniques to enhance the robustness of the models. The project's multifaceted objectives include a user-friendly interface, real-time object detection, multi-modal recognition, Text-to-Speech audio output, and an overarching aim of enriching the lives of visually impaired individuals. Driven by the global prevalence of visual impairment and the demand for cost-effective solutions, "Visually" is aligned with international efforts for accessibility and inclusivity. For cross-platform compatibility, the machine learning models have been integrated whilst being deployed with TensorFlow Lite. With Offline availability, the application ensures accessibility even in rural areas with limited network connectivity. To make a substantial societal impact "Visually" aims to contribute to a more inclusive and equitable society, by transforming the way visually impaired individuals navigate around the environment. Positioned at the intersection of technology, accessibility, and empowerment, the "Visually" project is poised to bring about positive change for a community that frequently encounters unique challenges in their daily lives.

Keywords: Visually Impaired, Assistive Technology, Mobility Aids, Navigation Assistance, Vision Impairment, AI-Assisted Mobility

Introduction:

In the realm of assistive technology, individuals with visual impairment face challenges every day that significantly impact their self-sufficiency, independence, and mobility. Despite the various existing technological aids, the challenge of navigating independently in unacquainted environments remains a pervasive issue, often requiring reliance on sighted assistance. This paper aims to address such challenges through the introduction of "Visually," an innovative mobile application developed to empower the visually impaired community. The primary goal of the project is to offer a real-time vision system, incorporating object detection, face recognition, and currency recognition features. These functionalities are integrated into the "Visually" application, incorporating deep learning models trained on a diverse dataset carefully curated for the unique needs of the visually impaired. This introduction outlines the background, research problem, and significance of the "Visually" project, which aims to ameliorate the daily experiences of visually impaired individuals by enhancing their independence, accessibility, and mobility. The main contributions of this research include addressing the challenges faced by visually impaired individuals, the global prevalence of visual impairment, and the development of an affordable and accessible solution with the potential to have a positive societal impact. The following sections will delve into the technical details of the model architectures, training methods, and the expected impact of the "Visually" application on its users.

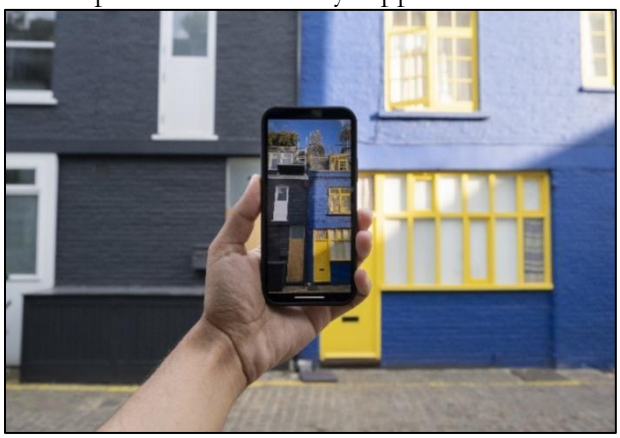

Figure 1: Navigation support using mobile application. [1]

Figure 1 illustrates the navigation support provided by the mobile application developed in the "Visually" project. The image shows a person using a smartphone application to navigate their surroundings, with the application providing real-time information about the user's environment through object detection and voice output.

Related Work:

Microsoft's "Seeing AI":

The landscape for assistive technologies for visually impaired individuals has witnessed significant contributions from notable projects and research endeavors. One of the major initiatives is Microsoft's "Seeing AI," a cloud-powered application that integrates Machine Learning (ML) to describe the user's surroundings and convert it into voice output, including features like text recognition, scenic descriptions, and more. "Seeing AI" can be considered a valuable benchmark for understanding the widespread usage beneficial for visually impaired users, influencing the development decisions and design considerations for the "Visual" application [2].

Notify (Currency Recognition):

Furthermore, "Notify", an application made in India for currency detection focuses on verifying accurate financial handling for visually impaired users, making it a worthy solution for the requirement [3].

Cutting-Edge Object Detection Techniques:

In the field of advanced object detection, Mahendran and his team explore cutting-edge solutions based on real-time vision systems using deep learning and point cloud processing. By investigating these techniques, the team helps the application recognize objects quickly and accurately. Their research is crucial for refining the goals of the "Visually" application, especially in computer vision and object recognition [4].

Deep-See Face Framework:

The DEEP-SEE FACE framework introduces an intriguing approach using convolutional neural networks (CNNs) for real-time facial recognition. This technology provides valuable insights for developing facial recognition models, focusing on features like hard negative mining and acoustic communication. These aspects align with "Visual's" goals of improving user understanding and communication, making the application more user-friendly and effective for visually impaired individuals [5].

Currency Recognition System:

A notable innovation is a currency recognition system designed for the visually impaired, enabling real-time identification of currency notes and obstacle-aware navigation. This system offers essential considerations for integrating an efficient and accurate currency recognition feature into the "Visually" application, ensuring that users can manage their finances independently and safely [6].

Summary:

In summary, the insights from these different sources serve as guiding principles for the development of the "Visually" application. Each project and research paper provides unique perspectives and considerations, collectively shaping innovative, user-centric solutions aimed at transforming the lives of visually impaired individuals. Through the integration of these technologies, "Visually" aims to empower users, enhancing their independence and quality of life.

Material and Methods:

The project's research design was meticulously crafted to ensure the effective implementation of its goals. The development of the "Visually" project follows a thorough and systematic methodology, covering all stages from data collection to model deployment. This approach is detailed in the following sections, providing a glimpse into the research design, data acquisition process, model training, and deployment strategies. Fig 2 demonstrates the development stages for the visual application.

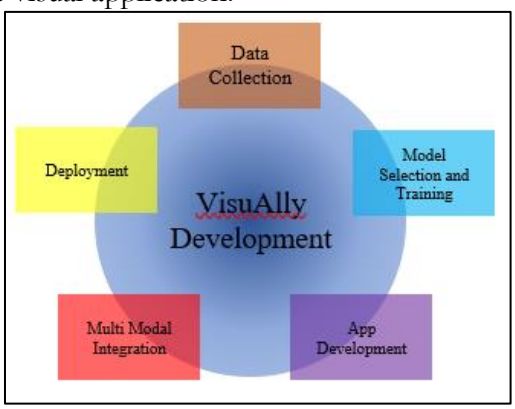

Figure 2: Development Stages for Visually App

Data Collection:

The "Visually" Project is built upon a rich and extensive dataset, specifically tailored and designed to meet the unique requirements of visually impaired people. The dataset includes a wide variety of indoor and outdoor objects, with a focus on objects essential for daily navigation. To ensure the model, that is to be trained on this dataset, is adaptable to various real-world

scenarios, the dataset undergoes careful augmentation and labeling. Techniques such as rotation, scaling, and flipping the images are employed to improve the model's ability to learn different features of the images and enhance its accuracy in identifying objects from different angles and lighting conditions.

Model Architecture and Training:

The core of the "Visually" application is its deep learning models, which utilize the "You Only Look Once (YOLO)" architecture known for its ability to detect objects in real-time videos with high accuracy. The YOLO model is pre-trained on a diverse dataset but can be self-trained on specific datasets to detect specific objects. This model is trained on the augmented dataset using advanced optimization methods, such as stochastic gradient descent. During training, the focus is on minimizing detection loss while enhancing the model's accuracy in recognizing and pinpointing objects in the user's environment.

Table 1: Differences between Yolo Architecture and Mobile Net SSD

Features	YOLOv5	Mobile Net SSD
Number of Layers	Backbone: 53 layers Additional: Depends on variant	Backbone: 13 layers Additional: 6 layers
Optimizer Used	Adam	Adam
Number of Epochs	Around 300	Around 200
Dataset	COCO dataset	VOC dataset
Images Used	Around 118K images	7,000 images
Average Accuracy	mAP of 50.1 on COCO test set	mAP of 72.7 on VOC test set
Frames per Second	60+ FPS on a GPU	Over 100 FPS on modern hardware
Model Size	27 MB	23 MB

Table 1 compares two popular object detection models, YOLOv5 and Mobile Net SSD, across various metrics. YOLOv5 is a state-of-the-art object detection model known for its efficiency and accuracy. It features a highly optimized architecture, enabling it to achieve real-time inference speeds on GPU hardware. YOLOv5 has gained popularity for its simplicity, ease of use, and impressive performance on a variety of datasets. It offers a range of pre-trained models of different sizes (e.g., YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x), allowing users to choose the right balance between speed and accuracy for their specific application. Additionally, YOLOv5 is actively maintained and updated, ensuring compatibility with the latest advancements in object detection research. Overall, YOLOv5 is a compelling choice for developers seeking a powerful and efficient object detection solution.

Multi-Modal Recognition:

The "Visually" application offers a comprehensive user experience by integrating various recognition features such as object, face, and currency recognition. Face recognition utilizes the weights of the pre-trained models and performs personalized training to achieve precise identification. Currency recognition involves the development of a specialized dataset and model, providing users with information about the currency notes they come across [7].

Text-to-Speech (TTS) Integration:

To enhance communication with visually impaired people, the "Visually" application provides an audio output feature powered by Text-to-Speech (TTS) technology. This feature not only describes the objects but also provides contextual information, such as the proximity to the obstacles, identifying which feature the user is currently using, and giving an audio output of the recognized faces, etc.

Deployment:

To achieve compatibility across Android and IOS platforms, all machine/deep learning models in the application are converted into TensorFlow Lite (tflite) versions. The application is developed using the Flutter framework, which utilizes the tflite_flutter_helper package for efficient image processing and input preprocessing.

Offline Functionality:

Recognizing the importance of accessibility in areas with limited network coverage, "Visually" emphasizes offline functionality. By using tflite files, all the models will be stored on the device which will allow the user to make use of the features even without internet services. Face embeddings and emergency contacts are stored locally, reducing the reliance on a constant internet connection. This approach aligns with the project's goal of creating a user-friendly, adaptive, and accessible solution for visually impaired individuals, ensuring that the application effectively addresses their unique challenges and contributes to positive societal impact [8].

Results and Discussion:

Table 2 provides a detailed breakdown of the accuracy achieved by the "Visual" application in detecting various objects essential for visually impaired individuals. The table showcases the impressive accuracy rates for detecting different objects, highlighting the application's proficiency in real-time object recognition. With a focus on key objects like people, laptops, water bottles, furniture, cars, motorcycles, and houseplants, the "Visually" application demonstrates high accuracy levels, ranging from 90% to 99%. These accuracy rates underscore the application's ability to swiftly and accurately identify a diverse range of objects, enhancing users' spatial awareness, safety, and overall navigation experience.

By utilizing the YOLO-based object detection model, "Visually" can swiftly and accurately identify a wide range of objects in real-time, such as obstacles, furniture, and other navigation aids. This feature offers users instant information about their environment, improving their spatial awareness and safety. Our thorough assessment of "Visual's" object detection abilities revealed exceptional performance, demonstrating the reliability and effectiveness of the integrated technologies.

Laptops Detection:

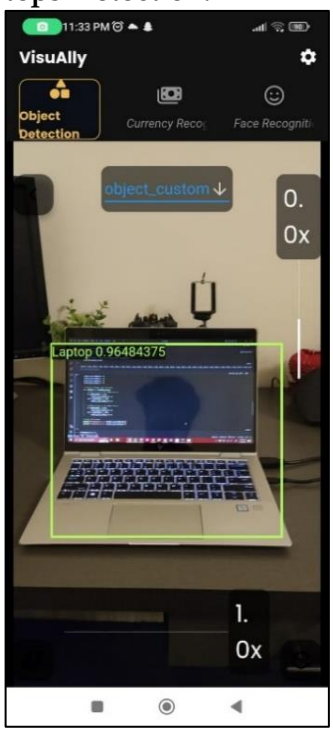

Figure 3: Laptop Detected

Figure 4: Bottle Detected

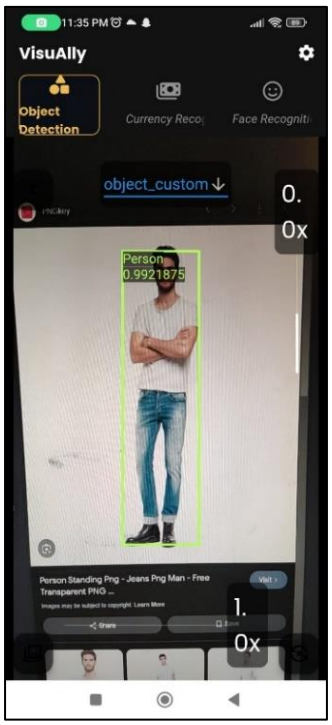

Figure 5: Person Detected

Figure 3 shows the snapshot of the application detecting a laptop. The model demonstrated exceptional proficiency in identifying laptops, achieving an impressive accuracy of 96%. This high level of accuracy is essential for users who rely on the application to navigate environments where laptops are prevalent.

Bottle Detection:

The model detected everyday objects, like water bottles, with exceptional accuracy, reaching 96%, as demonstrated in Fig 4. This ensures that the application can help a user in assisting with daily tasks such as locating personal items or identifying objects in their immediate vicinity.

Table 2: Accuracy achieved by Visually Application

Object Detected	Model Accuracy
Person	99%
Laptop	96%
Water Bottle	96%
Furniture	90%
Car	98%
Motorcycle	98%
Houseplants	96%

Person Detection:

Recognizing individuals in front of the user is a critical aspect of the application's functionality. In Fig 5, the detection of a person can be seen. Our model achieved an impressive accuracy of 99% in person detection, ensuring users are aware of the presence of others in their surroundings.

Car Detection:

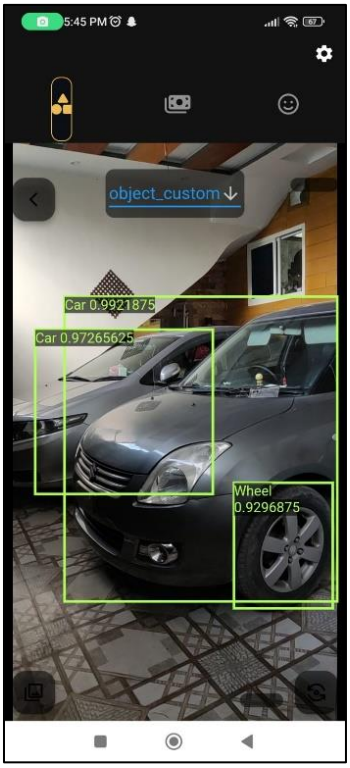

Figure 6: Car Detected

Figure 7: Multi Objects Detected

The most important objects to be detected in daily navigation are vehicles. Our model successfully detects cars in front of the user with a remarkable accuracy of 99%, making daily navigation easy for visually impaired people. Fig 6 demonstrates the snapshot of our application where it is detecting vehicles.

Multi-Object Detection:

The application's ability to detect multiple objects simultaneously further enhances its utility. With a focus on accuracy, the application provides users with detailed information about their environment, allowing for more informed navigation decisions. Fig 7 shows the multi-

object detection capabilities of Visually. Moreover, TTS API is utilized to announce each object to the user that is detected, providing detailed information about their environment and further enhancing the application's usability.

Discussion:

The "Visually" application enhances the mobility of visually impaired individuals through the integration of YOLO v5 for real-time object detection and Google's ML Kit for text-to-speech conversion. These technologies offer several key benefits:

Immediate Environmental Feedback:

By incorporating real-time object recognition technology, visually impaired people may navigate safely and autonomously by being promptly informed about their surroundings. Their improved spatial awareness makes them more competent navigators.

High Precision Object Identification:

The remarkable accuracy of YOLO v5 ensures accurate information about the surroundings is provided and informs the user about the items in their immediate environment, hence enhancing their spatial comprehension.

Efficient Navigation:

The application vocalizes the names of detected items, especially in unfamiliar locations, reducing users' cognitive load and allowing them to focus more on their surroundings.

Enhanced Safety:

The safety of visually impaired people is improved by accurate object detection and TTS announcements, which notify them of potential hazards or obstructions in their route. Because of the trustworthy information given to visually impaired people, they may navigate with more confidence thanks to this proactive approach to safety.

User-Friendly Interface:

The user-friendly interface of the application makes it easier even for consumers with little technological background. The incorporation of these technologies into a Flutter application results in an intuitive user interface that is simple to use. This accessibility is essential to guarantee that a broad spectrum of users, regardless of their level of technical expertise, can use the application efficiently.

In conclusion, the integration of YOLO v5 and Google's ML Kit in the "VisuAlly" application greatly enhances the mobility and independence of individuals with visual impairments. This integration allows them to confidently explore their surroundings, ensuring they have the necessary information for safe and effective navigation through the combination of real-time object detection and text-to-speech conversion.

Competitive Analysis:

“Visually” aims at providing an extensive solution to enhance the independence and mobility of visually impaired and blind people, exhibiting several significant advantages in comparison to the already available apps designed for the same objective. This discussion concentrates on elucidating the project’s results and implications, including the comparison of results with the available solutions, whilst highlighting the significant contribution of “Visually” to the field.

Comparison with Existing Apps:

“Visually” discerns itself from the multiple existing apps for visually impaired people through its encompassing approach and multi-modal features. Current apps mostly focus on singular functionalities such as object detection or navigation, whereas “Visually” incorporates real-time object detection with face recognition and currency recognition, making it an extensive solution catering to meeting the multiple needs of users. Additionally, features like offline mode, user-friendly design, and continuous user feedback integration make "Visually" unique [9].

Other popular apps like Be My Eyes and Seeing AI offer useful features such as remote assistance and text recognition. However, they may not offer the same depth of functionalities

and offline capabilities as "Visually." The use of YOLO architecture for object detection in "Visually" enhances its real-time processing capabilities, setting it apart from other solutions.

Significance of Results and Contribution to the Field:

The "Visually" project represents a significant advancement in assistive technology, specifically tailored to address the complex challenges faced by visually impaired individuals in their daily lives. Through its versatile and integrated approach, the application not only enhances users' independence but also reduces their reliance on others, fostering a sense of empowerment. One of the key strengths of "Visually" lies in its offline functionality, which ensures accessibility in a variety of environments, thereby filling a crucial gap in existing solutions. Moreover, the project's significance goes beyond its technical capabilities; it embodies a commitment to inclusivity, affordability, and user-centered design principles.

By aligning with global initiatives for accessibility, "Visually" acknowledges the widespread impact of visual impairment and strives to make a meaningful difference in the lives of affected individuals. The integration of user feedback further enhances the project's value, ensuring that the application evolves in response to the evolving needs of its users [10].

Conclusion:

"Visually" emerges as a groundbreaking assistive technology, designed to address daily challenges for visually impaired individuals. Utilizing advanced deep learning models, including real-time object detection, face recognition, and currency recognition, the application aims to transform the landscape of independence and mobility.

One of the key features of "Visually" is its commitment to a user-friendly interface and multi-modal recognition, including Text-to-Speech audio. This comprehensive approach is further enhanced by rigorous training on a diverse dataset, ensuring the application's adaptability to various real-world scenarios and reflecting its focus on meeting user needs. In line with global accessibility initiatives, "Visually" places a strong emphasis on affordability and offline functionality, which are essential for users in diverse environments. By combining the YOLO architecture for real-time processing with a holistic approach, the application surpasses existing solutions in its field.

What sets "Visually" apart from other apps is its offline capabilities, user-friendly design, and the integration of continuous user feedback. Beyond its technical capabilities, the application embodies inclusivity and user-centric design principles, contributing to a more equitable society. In conclusion, "Visually" represents a significant step towards empowerment for visually impaired individuals, redefining how they navigate the world and fostering inclusivity. With its responsive development cycle, the application is poised to have a lasting impact, evolving to meet the changing needs of its users and contributing to a more accessible world.

Acknowledgment: We acknowledge the usage of the GenAI platform for refining and enhancing the writing quality of this manuscript intended for submission at the ICTIS'24 conference.

References:

[1] "Seeing AI - Talking Camera for the Blind." Accessed: May 06, 2024. [Online]. Available: https://www.seeingai.com/

[2] "Noteify: Indian Currency Recognition App", [Online]. Available: https://github.com/chandran-jr/Noteify

[3] J. K. Mahendran, D. T. Barry, A. K. Nivedha, and S. M. Bhandarkar, "Computer vision-based assistance system for the visually impaired using mobile edge artificial intelligence," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., pp. 2418–2427, Jun. 2021, doi: 10.1109/CVPRW53098.2021.00274.

[4] B. Mocanu, R. Tapu, and T. Zaharia, "DEEP-SEE FACE: A Mobile Face Recognition System Dedicated to Visually Impaired People," IEEE Access, vol. 6, pp. 51975–51985, Sep. 2018, doi: 10.1109/ACCESS.2018.2870334.

[5] "Currency Recognition System Using Image Processing: Libyan Banknote as a Case Study ." Accessed: May 06, 2024. [Online]. Available: http://www.warse.org/IJETER/static/pdf/file/ijeter171022022.pdf

[6] R. C. Joshi, S. Yadav, M. K. Dutta, and C. M. Travieso-Gonzalez, "Efficient Multi-Object Detection and Smart Navigation Using Artificial Intelligence for Visually Impaired People," Entropy 2020, Vol. 22, Page 941, vol. 22, no. 9, p. 941, Aug. 2020, doi: 10.3390/E22090941.

[7] J. Madake, S. Bhatlawande, A. Solanke, and S. Shilaskar, "A Qualitative and Quantitative Analysis of Research in Mobility Technologies for Visually Impaired People," IEEE Access, vol. 11, pp. 82496–82520, 2023, doi: 10.1109/ACCESS.2023.3291074.

[8] "REAL TIME OBJECT DETECTION WITH SPEECH RECOGNITION USING TENSORFLOW LITE." Accessed: May 06, 2024. [Online]. Available: https://www.researchgate.net/publication/359393141_REAL_TIME_OBJECT_DETECTION_WITH_SPEECH_RECOGNITION_USING_TENSORFLOW_LITE

[9] N. Alsharabi, "Real-Time Object Detection Overview: Advancements, Challenges, and Applications," مجلة جامعة عمران , vol. 3, no. 6, pp. 12–12, Nov. 2023, doi: 10.59145/JAUST.V3I6.73.

[10] K. Vaishnavi, G. P. Reddy, T. B. Reddy, N. C. S. Iyengar, and S. Shaik, "Real-time Object Detection Using Deep Learning," J. Adv. Math. Comput. Sci., vol. 38, no. 8, pp. 24–32, Jun. 2023, doi: 10.9734/JAMCS/2023/V38I81787.