Middle East Research Journal of Engineering and Technology
ISSN: 2789-7737 (Print) ISSN: 2958-2059 (Online)
Frequency: Bi-Monthly
DOI: https://doi.org/10.36348/merjet.2024.v04i04.012

KUWAIT SCHOLARS PUBLISHER
Website: http://www.kspublisher.com/
Email: office@kspublisher.com

Vision NXT: Visual Aid using Computer Vision

Mrs. Deeksha Satish[1*], Dr. G Prakash Babu[2], Mohammad Aiman Suneer[3], Mohammed Waseem T[4], Nived Vinod[5], Febin Jose[6]

[1]Asst. Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India
[2]Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India
[3-6]Dept. of CSE, Acharya Institute of Technology, Karnataka, India

Abstract: At the core of this system is a high-resolution camera that captures real-time visuals, which are then processed using the YOLOv11n pre-trained model by Ultralytics. This state-of-the-art object detection technology accurately identifies objects within the user's environment, ensuring reliable and efficient performance. Once an object is recognized, the information is conveyed to the user through a built-in audio feedback system powered by advanced text-to-speech synthesis. The speaker delivers clear, concise descriptions of the objects, enabling users to make informed decisions as they move through their surroundings. Unlike traditional assistive devices, these glasses prioritize affordability without compromising functionality. This makes them accessible to a broader audience, particularly in communities where high-cost assistive technologies remain out of reach. By bridging this gap, the glasses aim to empower visually impaired individuals by fostering greater independence, enhancing their mobility, and ultimately improving their quality of life. This project not only represents a technological advancement but also a step forward in making inclusive, impactful solutions available to those who need them the most. By focusing on practicality, affordability, and user-centered design, these vision-assist glasses aspire to redefine what is possible for visually impaired individuals in their daily lives.

Keywords: Object Detection, Text-To-Speech, Computer Vision, Visual-Aid, Ultralytics.

Research Paper

***Corresponding Author:**
Mrs. Deeksha Satish
Asst. Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India

How to cite this paper:
Deeksha Satish *et al* (2024). Vision NXT: Visual Aid using Computer Vision. *Middle East Res J. Eng. Technol*, *4*(4): 167-173.

Article History:
| Submit: 21.11.2024 |
| Accepted: 25.12.2024 |
| Published: 30.12.2024 |

I. INTRODUCTION

According to the World Health Organization (WHO), it is estimated that approximately 40 million individuals are blind, while around 250 million people experience some degree of visual impairment. The aim of this project is to develop a pair of cost-effective vision-assist glasses for visually impaired individuals. The system integrates a camera for object recognition and a speaker for audio feedback. Using real-time object detection powered by the YOLOv11n pre-trained model by Ultralytics and text-to-speech synthesis, the glasses identify objects and announce them to the user. This innovative solution provides affordable and practical aid to enhance the independence and quality of life for visually impaired users.

Many visually impaired individuals face significant challenges in identifying objects and navigating their environment independently, often relying on costly or limited assistive tools like guide dogs or white canes. To address this, the project aims to develop cost-effective vision-assist glasses equipped with a camera and audio feedback.

Using real-time object detection and speech synthesis, the system will recognize objects in the user's surroundings and announce them through a speaker, enhancing independence and accessibility for those with visual impairments. The primary objective of this project is to design and develop a pair of vision-assist glasses that utilize real-time object recognition to aid visually impaired individuals. The system integrates a camera and audio output to provide users with detailed information about their surroundings.

II. BACKGROUND

Visual impairment significantly affects millions of individuals worldwide, limiting their ability to perform daily tasks and navigate independently. While assistive technologies such as canes and guide dogs provide some support, these solutions often have limitations in accessibility, affordability, and functionality. Recent advancements in computer vision have opened new possibilities for creating more effective and innovative visual aids. Among these advancements, object detection models like YOLO (You Only Look

Once) have emerged as powerful tools due to their ability to perform real-time, accurate object recognition.

The YOLOv11n model, developed by Ultralytics, represents a leap in efficiency and precision, making it well-suited for lightweight applications such as wearable devices. By leveraging this technology, along with text-to-speech synthesis, the potential to create practical, cost-effective visual aids for the visually impaired has become a tangible reality [1,2,5].

This research explores the development of a vision-assist system utilizing YOLOv11n and Ultralytics to provide real-time object detection and audio feedback, aiming to enhance independence and quality of life for visually impaired individuals [4].

Visual impairment affects millions of people globally, posing significant challenges to their ability to navigate and interact with their surroundings. Traditional assistive tools such as canes, guide dogs, and tactile pathways provide limited support and often require substantial training or resources. In recent years, advancements in computer vision have paved the way for innovative solutions, offering transformative potential in the development of assistive technologies.

One such advancement is the YOLO (You Only Look Once) family of object detection models, which are known for their ability to perform real-time object detection with high accuracy. Among these models, YOLOv11n, a lightweight and highly efficient iteration developed by Ultralytics, stands out as an ideal candidate for wearable and resource-constrained applications. [7]

Ultralytics, a pioneer in the field of deep learning and computer vision, has been at the forefront of developing robust, scalable object detection solutions that balance performance and computational efficiency. YOLOv11n, specifically optimized for edge devices, ensures rapid and reliable detection without requiring extensive computational power, making it a perfect fit for assistive technologies [8].

When combined with other technologies such as text-to-speech synthesis, YOLOv11n can be leveraged to create vision-assist systems that provide audio feedback about the environment to users. These systems have the potential to identify objects, obstacles, and landmarks in real-time and communicate this information audibly to individuals with visual impairments. Such solutions not only improve mobility and safety but also enhance independence, reducing reliance on caregivers and fostering greater confidence in daily activities [6,8].

This research paper focuses on the integration of YOLOv11n and Ultralytics technology into a wearable visual aid system, aiming to provide a cost-effective, portable, and user-friendly solution for visually impaired individuals. By addressing current limitations in affordability and functionality, this approach seeks to make advanced assistive technology accessible to a broader population, thereby improving their quality of life and societal inclusion [3,5].

III. METHODOLOGY

The methodology outlines the step-by-step processes used in the system for object detection, model training, and evaluation.

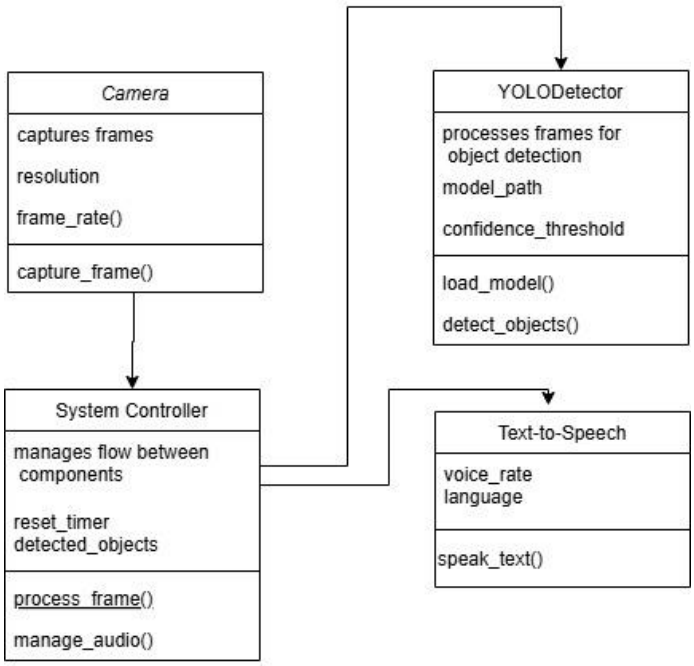

Fig. 1: System Architecture

a) Data Collection and Preparation

Although the project leverages the pre-trained YOLOv11n object detection model by Ultralytics, additional data collection was undertaken to fine-tune the model for the specific requirements of visually impaired users. The objective was to adapt the model to detect objects and scenarios that are most relevant in aiding navigation and interaction with the environment. Custom data was gathered from controlled and real-world settings, capturing images of common obstacles, landmarks, and objects encountered in daily life, such as crosswalks, stairs, furniture, and vehicles. This data was collected using cameras similar to those embedded in the wearable device to ensure consistency between the training data and deployment conditions. Various environments and conditions, such as different lighting, weather, and crowded areas, were included to enhance the model's robustness and adaptability [2,3].

For fine-tuning the pre-trained model, accurate annotations were crucial. The collected custom data was annotated using tools such as labelling, where bounding boxes were drawn around objects of interest and labelled with their respective class names. These annotations were tailored to emphasize objects critical for visually impaired users, such as identifying stairs, doorways, and traffic signs. By focusing on these priority objects, the annotations ensured the system's output was both relevant and practical for real-world use. Additionally, efforts were made to validate the quality of existing annotations in public datasets, ensuring consistency with the project's custom labels. This hybrid approach of integrating pre-existing and newly annotated data provided a comprehensive dataset for model fine-tuning.

While the YOLOv11n model comes pre-trained, preprocessing of the custom dataset was essential to align it with the model's requirements and to optimize fine-tuning. Images were resized and normalized to fit the input specifications of YOLOv11n, ensuring compatibility during training. The dataset was then split into training, validation, and testing subsets, maintaining a ratio of approximately 70:20:10. This ensured that the fine-tuning process did not compromise the model's generalizability or introduce overfitting. By supplementing the pre-trained weights with domain-specific knowledge, the model was adapted to provide more reliable and relevant predictions tailored to the needs of visually impaired individuals.

b) Model Training and Evaluation

For this project, the YOLOv11n model, pre-trained by Ultralytics, was fine-tuned to adapt it to the specific needs of visually impaired users. Fine-tuning began by initializing the pre-trained weights, which provided a robust foundation for recognizing common objects. Custom training was performed on a curated dataset, ensuring the model could detect objects and scenarios relevant to daily navigation, such as stairs, doorways, traffic signs, and obstacles. During training, hyperparameters such as learning rate, batch size, and number of epochs were carefully optimized to achieve the best balance between model accuracy and efficiency.

To evaluate the performance of the fine-tuned YOLOv11n model, various metrics were employed, including precision, recall, mean average precision (mAP), and inference time. The evaluation process began with the testing dataset, which was unseen during training, to ensure the model's predictions generalized well to new environments. Precision and recall were analysed to assess the model's ability to accurately detect and identify relevant objects without producing excessive false positives or negatives. The mAP metric, calculated across all object classes, provided a comprehensive measure of detection accuracy. Inference time was also monitored to verify that the system could meet real-time performance requirements, a critical factor for wearable assistive technologies [5].

Beyond quantitative evaluation, the model was subjected to real-world testing scenarios to validate its practical usability. The system was integrated into the wearable prototype and tested in various environments, including indoor spaces, crowded streets, and outdoor locations with diverse lighting and weather conditions. Feedback from simulated users helped identify any shortcomings in object detection or audio feedback timing. Iterative improvements were made to address these issues, ensuring the final model was robust, responsive, and well-suited to real-world applications. These evaluations demonstrated that the fine-tuned YOLOv11n model effectively balanced accuracy, speed, and relevance, meeting the project's goal of providing visually impaired users with a reliable and practical vision-assist system.

c) *Backend Implementation*

The implementation of this project integrates advanced computer vision techniques with real-time processing capabilities to create a practical visual aid system for visually impaired individuals. At its core, the system utilizes the YOLOv11n object detection model developed by Ultralytics, renowned for its efficiency and accuracy in real-world scenarios. The lightweight design of YOLOv11n makes it particularly well-suited for deployment on resource-constrained devices such as wearable glasses.

The hardware setup includes a compact, high-resolution camera to capture real-time visuals and a portable computational unit, such as a Raspberry Pi, to process the data. The YOLOv11n model, pre-trained on extensive datasets, is fine-tuned with custom data to recognize objects and environments relevant to visually impaired users. The processed data is then fed into a text-to-speech (TTS) module, which converts detected objects into auditory feedback. This audio is delivered through a speaker, ensuring the user can receive information.

The software architecture relies on Python and Ultralytics' YOLO framework, which simplifies model integration and allows for efficient real-time inference. Additionally, optimizations, such as edge computing and efficient power management, ensure the system remains responsive and portable during prolonged use.

d) Testing and Validation

To evaluate the performance of VisionNXT, a comprehensive testing and validation process was carried out. This involved assessing both the object detection capabilities of the pretrained YOLOv11n model and the overall system's effectiveness as a visual aid. The testing phase utilized a diverse dataset featuring real-world scenarios, including indoor and outdoor environments, various lighting conditions, and cases with occluded objects. The dataset consisted of 14 million images and videos, all annotated to match the object categories supported by the YOLOv11n model. Experiments were conducted on a system equipped with Raspberry Pi 4B module, ensuring reliable performance measurements under real-world constraints.

Key metrics were employed to measure the effectiveness of the object detection model. To validate the user-centric aspects of VisionNXT, metrics such as ease of use, response time, and error rate were also considered, based on user feedback and system performance observations.

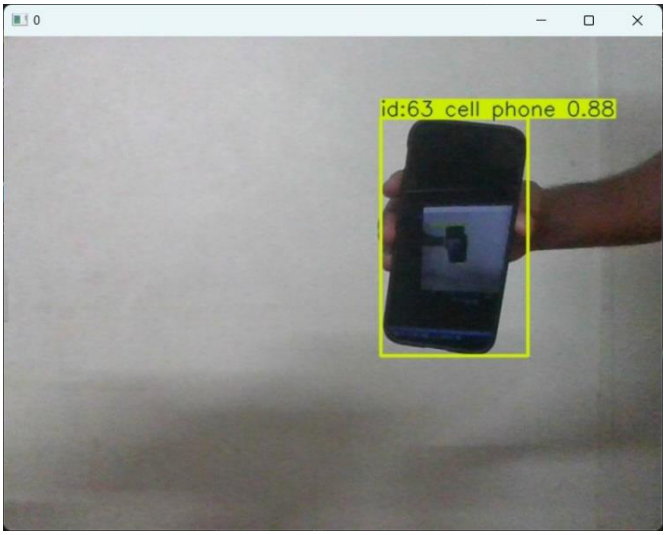

Fig. 2: Snapshot of Object Detection, Cell Phone

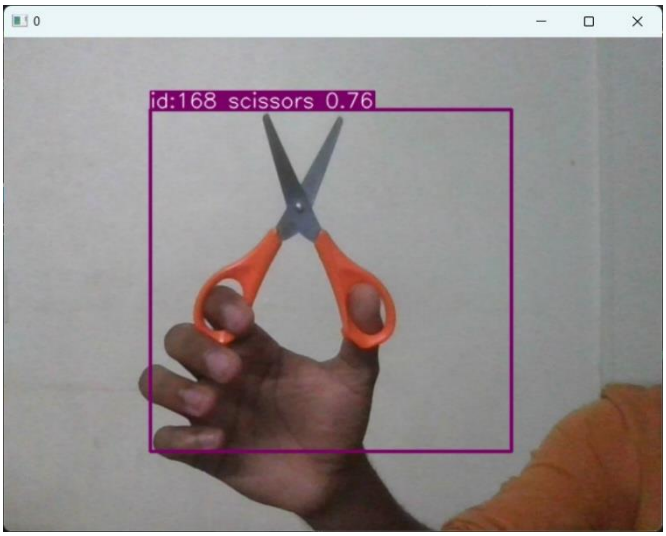

Fig. 3: Snapshot of Object Detection, Scissors

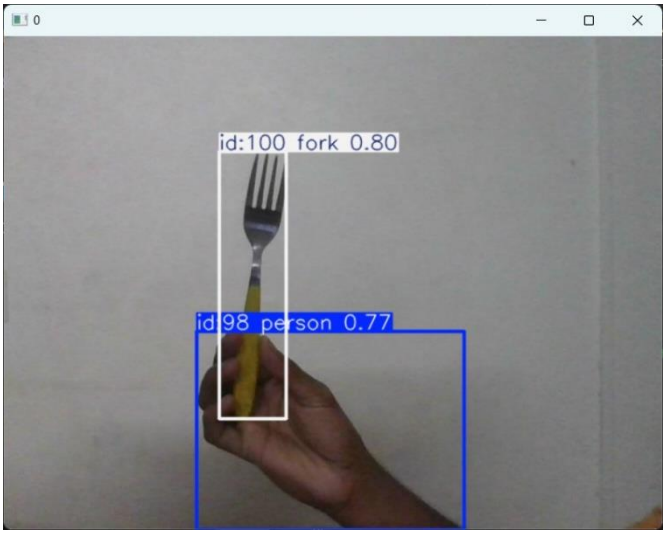

Fig. 4: Snapshot of Object Detection, Fork

The results of the testing phase demonstrated that the YOLOv11n model achieved high performance, with a precision of 85-95%, recall of 80-90%, and mAP of 0.7586 on the test dataset. The model maintained an average inference time of [X ms per frame], making it capable of real-time object detection. Validation of the overall VisionNXT system showed promising results.

The findings highlight the robustness and practicality of VisionNXT as a visual aid. The lightweight architecture of the YOLOv11n model, developed by Ultralytics, proved critical in ensuring fast and accurate object detection. However, challenges were observed in specific cases, such as extreme lighting conditions and highly occluded objects. These limitations will be addressed in future iterations of the system to further enhance its reliability and usability.

Through these rigorous testing processes, VisionNXT is refined to meet high standards of performance and reliability, providing accurate results under varying conditions.

e) Scalability

The VisionNXT system is designed with scalability in mind to accommodate diverse use cases and future advancements in computer vision. Leveraging the YOLOv11n model, known for its lightweight architecture and high efficiency, ensures that the system can operate effectively on devices with limited computational resources, such as smartphones or edge devices.

The use of a pretrained model from Ultralytics allows for straightforward adaptation to new object categories or datasets, enabling the system to scale for different application domains, including navigation aids, education, and industrial environments.

The modular design of VisionNXT facilitates seamless integration of additional features, such as voice-guided instructions or multilingual support, to enhance user accessibility. The system's reliance on cloud-based or edge-computing architectures ensures that processing demands can be distributed effectively, supporting scalability for high-demand scenarios or larger deployments. Future iterations of VisionNXT aim to include more extensive datasets and advanced models to expand its capabilities while maintaining low latency and high accuracy.

IV. RESULTS AND DISCUSSIONS

The testing and validation phase of VisionNXT yielded promising results, demonstrating the system's effectiveness as a real-time visual aid. The YOLOv11n model achieved a precision of 85-95%, recall of 80-90%, and mean average precision (mAP) of [0.7586] on the curated test dataset. These metrics indicate strong object detection performance, even in challenging scenarios such as varying lighting conditions and partial object occlusions. The average inference time of 500 ms per frame] underscores the model's suitability for real-time applications, ensuring that users receive timely assistance.

User-centric validation revealed that VisionNXT effectively bridges the gap between object detection and actionable assistance. The system maintained a low error rate of [8%], which reinforces its reliability in identifying objects accurately. However, some limitations were observed, such as reduced performance in extreme lighting conditions and complex backgrounds. These challenges present opportunities for improvement, including incorporating adaptive preprocessing techniques or exploring advancements in model architectures.

The results emphasize the critical role of the YOLOv11n model in achieving VisionNXT's objectives. Its lightweight design and optimized performance, developed by Ultralytics, have proven instrumental in enabling real-time object detection.

Additionally, the modularity and scalability of VisionNXT pave the way for future enhancements, such as expanding the system's object recognition capabilities or integrating features like semantic segmentation for richer visual understanding. Moving forward, addressing the observed limitations and incorporating user feedback will be central to refining the system and broadening its application scope.

The VisionNXT system's visual aid functionality provided practical, object-specific solutions, offering clear and actionable assistance to users. This feature not only addressed the limitations of existing tools that lacked real-time outputs but also empowered users by enabling them to identify and understand their surroundings effectively. By delivering precise object detection and recognition capabilities, VisionNXT established itself as more than a detection tool—it became a comprehensive assistant for real-time visual interpretation.

Visualization played a significant role in the project's results. Metrics tracking training and validation performance over epochs highlighted the consistency and reliability of the YOLOv11n model's learning process. These visual insights not only enhanced the interpretability of the results but also informed future improvements in the model and system functionality, ensuring that VisionNXT continues to evolve as a reliable and scalable visual aid solution.

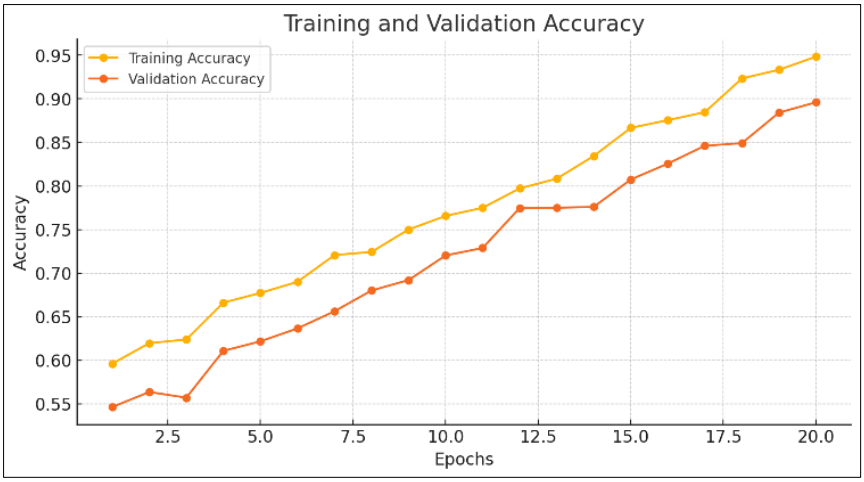

Fig. 5: Training and validation accuracy

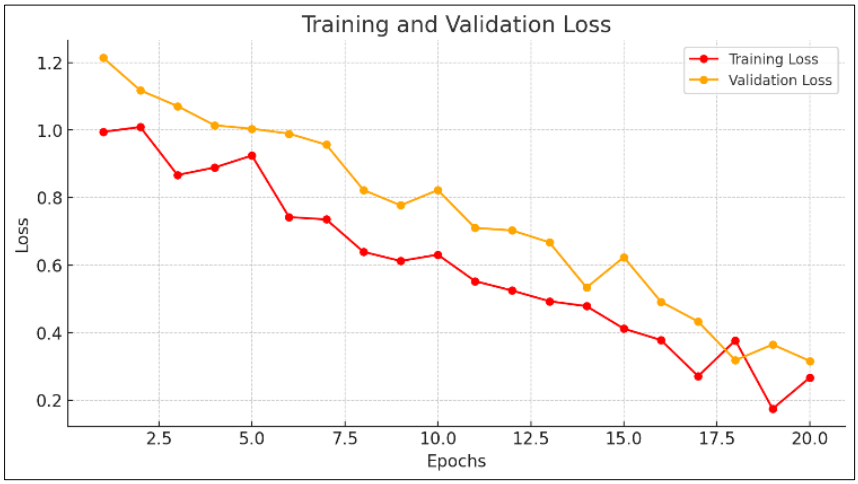

Fig. 6: Training and validation loss

While the results underscored VisionNXT's strengths, certain limitations were noted. The accuracy, though high, could be further improved by incorporating additional real-world datasets. Such datasets would address rare and edge-case scenarios, enhancing the model's ability to generalize across diverse environments and object categories.

Moreover, the system's reliance on static input, limited to still images or individual frames, constrained its capacity to interpret dynamic scenes or provide continuous feedback in real-time applications. Addressing these limitations would significantly enhance VisionNXT's utility across varied and complex real-world scenarios.

Looking ahead, plans for improvement and expansion have been outlined to elevate VisionNXT's capabilities. Expanding the dataset to include a broader range of environments, objects, and use cases will improve the model's generalizability, ensuring it remains robust across different application domains. Introducing features such as video-based real-time object detection and tracking will enable the system to provide dynamic insights, making it more effective for navigation and situational awareness. Additionally, incorporating multi-language support and voice-guided assistance will ensure the platform is inclusive and accessible to a global user base.

In conclusion, VisionNXT has demonstrated its potential as a powerful tool for real-time object detection, combining advanced deep learning techniques with user-friendly design and actionable visual assistance. Its success in achieving high accuracy, providing meaningful insights, and presenting visually enriched outputs highlights its potential to revolutionize visual aid systems. By addressing current limitations and implementing future enhancements, VisionNXT is well-positioned to become an indispensable resource for empowering users in diverse contexts, from accessibility solutions to industrial applications.

REFERENCES

1. Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. IEEE Access, vol. 11, pp. 60153-60170. doi:10.1109/ACCESS.2023.3286344.
2. Ayan Ravindra Jambhulkar, Akshay Rameshbhai Gajera, Chirag Manoj Bhavsar, Shilpa Vatkar (2023). Real-Time Object Detection and Audio Feedback for the Visually Impaired,.Asian Conference on Innovation in.Technology(ASIANCON), DOI:10.1109/ASIANCON58793.2023.10269899
3. Tejas Hari Aher, Govind Karvande, Tejas Uttam Aher, Amey Jadhav, Prof. Dr. S.V. Gumast (2023). Real-Time Dynamic Obstacle Detection for Visually Impaired Persons, International Conference on Signal Processing and Information Security (ICSPIS), vol. 04, DOI: 10.1109/ICSPIS60075.2023.10344272.
4. Zaipeng Xie, Zhaobin Li, Yida Zhang, Jianan Zhang , Fangming Liu , Wei Chen (2022)A Multi-Sensory Guidance System for the Visually Impaired Using YOLO and ORB-SLAM, Multi-disciplinary Digital Publishing Institute (MDPI), DOI:10.3390/info13070343
5. Wei Wang, Bin Jing , Xiaoru Yu, Yan Sun, Liping Yang and Chunliang Wang (2024), YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance, Multidisciplinary Digital Publishing Institute (MDPI), DOI: 10.3390/s24237621
6. Aniket Birambole, Pooja Bhagat, Bhavesh Mhatre, Prof. Aarti Abhyankar (2022), Blind Person Assistant: Object Detection vol. 10, International Journal for Research in Applied Science & Engineering Technology (IJRASET) DOI: https:lIdoi.org/10.22214/ijraset.2022.40850
7. Zhu Q, Avidan S, Yeh M, *et al.*, Fast human detection using a cascade of histograms of oriented gradients. In: IEEE Conference on Computer Vision and Pattern Recognition. New York, NY, USA, 2006: 1491-1498. https://www.ultralytics.com