International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2410232

VI-Assist Using AI for Visually Impaired Person

Riyanshu Rai, Adil Khan, Neha Singh, Dr. Vinayak Shinde, Ashish Pal

Department of Computer Engineering, Shree L. R. Tiwari College of Engineering, Mumbai, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 20 March 2024

Published: 03 April 2024

Publication Issue

Volume 10, Issue 2

March-April-2024

Page Number

378-387

ABSTRACT

Vi-Assist is a ground-breaking tool that offers a wide range of capabilities to meet the various issues faced by people with visual impairments. Utilizing state-of-the-art technologies like YOLOv5 for object detection, BLIP for environment description, and an advanced path navigation algorithm based on A*, the app offers real-time information, enabling users to navigate, interact with their surroundings, and find objects of interest more effectively. Furthermore, Vi-Assist uses Deep Face for facial recognition, supporting users in recognizing known faces and deciphering non-verbal signs to overcome obstacles in social interactions. MIDAS for depth estimation, OpenCV, Deep Learning, PyQt, AI/ML techniques, and Eleven Labs for AI speech synthesis are all integrated into this revolutionary application, which goes beyond simple assistance to empower visually impaired people and promote confidence, independence, and enhanced standard of living overall.

Keywords — Vi-Assist, Object Detection, Path Navigation Algorithm, Depth Estimation, AI Speech Synthesis.

I. INTRODUCTION

Innovative solutions are required to empower those who face the primary causes of visual impairment, which impact 43% and 33% of people worldwide, respectively, and include uncorrected refractive problems and cataracts [1]. Vi-Assist stands at the forefront of accessibility technology, offering a groundbreaking solution to the myriad challenges faced by individuals with visual impairments [2]. This revolutionary application seamlessly integrates cutting-edge technologies, employing YOLOv5 for real-time object detection , BLIP for detailed environment description, and an advanced A*-based path navigation algorithm . By leveraging these state-of-the-art features, Vi-Assist provides users with crucial information and support for efficient navigation through various surroundings. The incorporation of Deep Face for facial recognition and MIDAS for depth estimation enhances the app's capabilities, empowering visually impaired individuals to recognize familiar faces and perceive the spatial layout of their environment . Furthermore, Vi-Assist goes beyond conventional accessibility tools with its integration of OpenCV, Deep Learning, PyQt, and Eleven Labs for AI speech synthesis . This comprehensive approach ensures a user-friendly interface and advanced image processing capabilities, allowing for a more enriched and interactive user experience. The overarching goal of Vi-Assist is not just assistance but empowerment. The application strives to elevate the standard of living for visually impaired individuals by promoting confidence, independence, and a sense of self-assurance through its multifaceted features [3] . In summary, Vi-Assist stands as a testament to the transformative power of technology, fostering inclusivity and enhancing the lives of those with visual impairments through its innovative and holistic approach [4].

II. RELATED WORK

F. Ashiq et al. in [4] The article presents a high-tech system intended to improve visually impaired people's (VIPs) mobility and safety. The system, which makes use of a deep convolutional neural network (CNN) model, performs exceptionally well in real-time item identification and recognition, which enhances VIPs' ambient awareness. Real-time support is provided through the seamless integration of automated voice instruction. Interestingly, the system's usefulness is expanded by a web service that lets VIPs' family follow their whereabouts and request pictures whenever they want. Based on increased accuracy and sophisticated features, the system is said to operate better than current devices. Through the seamless integration of state-of-the-art technology, on-demand assistance, and family connectivity, this intelligent solution offers a potential path toward enhancing the safety and independence of visually impaired people in their daily lives. Furthermost, The small, unrepresentative sample size that affects generalizability, the unclear data collection methods that raise questions about data quality, the superficial analysis that ignores confounding factors that undermine causal claims, and the absence of a clear theoretical framework that makes it difficult to contextualize and evaluate the paper's contribution are some of its limitations. For increased scholarly rigor, significant advancements in design, methodology, analysis, and writing are required.

M. A. Khan et al. in [5] The aforementioned research presents a novel visual aid designed for people who are completely blind. This system uses a Raspberry Pi 3 Model B+, a camera, sensors, and image processing algorithms to provide functions including object detection, obstacle avoidance, and reading assistance. The design is noteworthy for its emphasis on price, compactness, and easy integration with regular spectacles. The device, according to the report, promises a potential development beyond traditional tools like the classic white cane and attempts to improve accessibility, comfort, and ease of navigating for the visually impaired. Although creative, the paper's visual help approach has some serious drawbacks. Its accuracy may be compromised due to its reliance on a single camera and lack of real-world testing. The simple text converter has problems with complicated content, and the fixed focal length camera may not perform well in different lighting conditions. When using earphones with auditory input, users may experience discomfort and interference, particularly in noisy surroundings. For the system to be useful for visually challenged people, these problems must be fixed.

Amit, Y., Felzenszwalb, P., Girshick, R. In [6] One of the core tasks in computer vision is object detection, which is locating and classifying things in pictures or videos. Applications for this adaptable method can be found in many different fields, such as robots, autonomous driving, security, surveillance, and medical imaging. There are two main types of object detection techniques that come into play: conventional techniques and deep learning techniques. While deep learning approaches use neural networks to automatically learn and infer complicated patterns, traditional methods frequently rely on handmade features and specialized algorithms. This makes deep learning methods more accurate and versatile for object detection. As technology develops, object detection systems' accuracy and efficiency have significantly increased with the incorporation of deep learning techniques, which has led to their widespread use in a variety of real-world applications. Object detection is still a field of current research with prospects and challenges. These days, the main goals are to improve efficiency and accuracy by using new architectures, loss functions, and data augmentation. Taking on challenges like as scale variation and occlusion is important, and combining attention and context enhances object detection. Furthermore, a crucial area of research is merging object detection with associated tasks such as segmentation and tracking.

C. Liu et al. In [7] suggests an object identification technique based on the YOLO network, a deep learning model with real-time object detection capabilities. The research focuses on how real-world image noise, blur, and rotation jitter impact object detection's robustness and accuracy. In order to replicate the actual shooting settings, the article creates image degradation models using traffic signs as an example. The average precision (AP) of traffic sign detection is then increased by training a robust model based on the YOLO network and comparing the effects of various degradation models on object detection. The study has flaws in that it doesn't provide a thorough performance comparison between the Multi-Head Attention Transformer and other text classification techniques. Unexplored are the ethical and societal ramifications, including prejudices and false information. Unaddressed are the issues of scalability, generalizability, and interpretability in various news environments. Additionally, there is no evaluation of user input or satisfaction with the news suggestion mechanism.

Wayahd et al. In [8] compares the A-Star, Dijkstra, and Greedy methods for determining a graph's shortest path. The study tests the algorithms' performance on a variety of network types, including sparse, dense, random, and grid graphs, using a simulation-based methodology. The number of nodes visited, the execution time, and the accuracy of the solution are the three metrics used in the study to measure performance. According to the paper, the Dijkstra's method is optimal but slow, the A-Star algorithm performs better than the Greedy algorithm but requires complicated data, and the Greedy algorithm is quick but not ideal. The paper does not consider the impact of the graph size, the edge weight, and the source-destination pair on the performance of the algorithms, which may limit the generalizability and the applicability of the findings

Bhoi A. In [9] describe monocular depth estimation: given a single RGB image as input, the aim is to forecast the depth value of each pixel of an image. The study examines five studies that address this issue using various methods, including supervised, weakly-supervised, and unsupervised learning. Based on their approaches, datasets, assessment metrics, and outcomes, the publications are compared in this paper. Future directions and problems for monocular depth estimation are also covered in this work. The paper does not consider the recent advances and developments in monocular depth estimation, such as self-attention, multi-task learning, or geometric constraints.

S. Li and W. Deng In [10] Provides a Recent developments in deep facial expression recognition (FER) are thoroughly examined in this review, which covers algorithm designs, datasets, protocols, and application scenarios. It provides an overview of changing network designs and loss functions, tackling important problems such as expression-unrelated variations and overfitting caused by inadequate training data. The document provides guidance for data selection and evaluation standards and lists commonly used datasets. It reviews the state of the art, compares the results on benchmarks, and assesses new deep neural networks and training methods for both static images and dynamic sequences. The survey covers affective computing, facial action unit identification, and cross-database and cross-cultural FER, offering a comprehensive view of the field's advancements and difficulties. The lack of a theoretical examination in the research about the complexity and accuracy of the algorithms could potentially compromise the validity and trustworthiness of the results.

The survey aimed to investigate visual impaired assistance from 2010 to 2023, examining a range of research and technological advancements designed to enhance the lives of individuals with visual impairments. This comprehensive review covers various aspects, including assistive devices, navigation systems, object detection technologies, and communication aids, providing insights into the evolution and current state of the field. The surveyed literature encompasses innovative solutions, methodologies, and frameworks that contribute to improving accessibility, independence, and overall well-being for individuals with visual impairments during the specified timeframe.

Paper	Focus	Methodology	Strengtl
[4]	Visual Aid for VIPs	High-tech system, CNN model	Real-tin identific Automa support, connect
[5]	Visual Aid for Blind	Raspberry Pi 3 Model B+, Camera, Sensors	Object Obstacl avoidan Reading
[6]	Object Detection in Computer Vision	Deep learning vs. Traditional methods	Increase accuracy efficienc Versatil
[7]	YOLO Network for Object Identification	YOLO network, Real-world image degradation models	Real-tin detectio Robustn accuracy assessme
[8]	Pathfinding Algorithms	A-Star, Dijkstra, Greedy methods	Perform metrics, Simulat method
[9]	Monocular Depth Estimation	Supervised, Weakly-supervised, Unsupervised learning	Compar approac Datasets directio
[10]	Deep Facial Expression Recognition	Algorithm designs, Datasets, Application scenarios	Overvie network Loss State-of review

Table 1. Comparative analysis of literature

Table 1. refers to various papers on topics like visual aid, object detection, pathfinding, depth estimation, and facial expression recognition, highlighting their methodologies, strengths, and weaknesses.

III. METHODOLOGY

Fig1 illustrates, The system involves crucial phases starting with data collection from a database and user input via a camera to build a diverse image dataset. Preprocessing includes aligning, cropping, and resizing images for consistency. The main model is trained and tested using various algorithms like YOLO, Midas, and OpenCV for object detection, navigation, and emotion recognition. Testing results guide predictions, providing outputs for object detection, navigation, and emotion recognition, ultimately generating a response for the user.

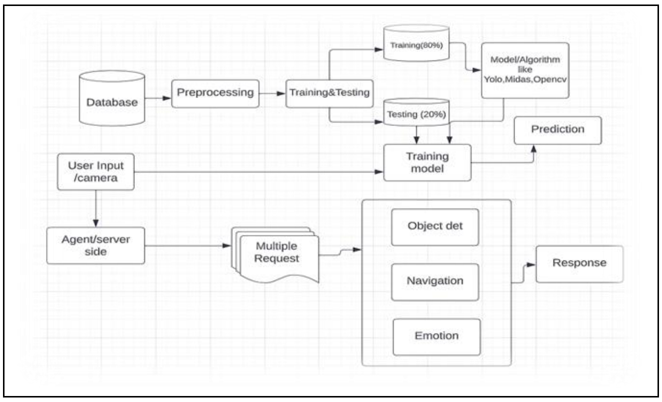

Fig1 .System Architecture

A. Data Collection

Object detection is the task of locating and identifying objects of interest in images or videos. Data collection is an important step for training and evaluating object detection models, as it provides the input images and the ground truth labels. There are many challenges and issues related to data collection for object detection, such as data quality, data quantity, data diversity, data annotation, data augmentation, and data privacy [11].

B. Preprocessing

Preprocessing readies input images for object detection, enhancing quality and reducing computational complexity. It addresses challenges like occlusion and scale variation, optimizing system efficiency for accurate navigation in diverse visual scenarios [12].

C. Algorithm and Model

Vi-Assist makes use of cutting edge technology, such as the A* algorithm for path navigation, BLIP for environment description, and YOLOv5 for object detection. In order to offer a comprehensive and efficient solution for the visually challenged, it also includes Deep Face for facial recognition, MIDAS for depth estimation, OpenCV for computer vision, and AI/ML approaches. Together, these models and algorithms provide real-time information processing, giving users the ability to navigate, interact with objects, recognize faces, and perceive depth.

1. YOLO v5: YOLOv5 is a model within the You Only Look Once (YOLO) family of computer vision models, known for detecting objects in images or videos. It offers four versions: small (s), medium (m), large (l), and extra-large (x), with varying accuracy and training times. YOLOv5 employs a convolutional neural network (CNN) that divides input images into a grid of cells. Each cell predicts bounding boxes, confidence scores, and class probabilities, facilitating fast and accurate object detection. YOLOv5 is recognized for its efficiency, precision, and user-friendly design.

2. Midas: A depth estimation model called MiDaS (Mixed Data Sampling) is intended to forecast an image's depth information. To estimate each pixel's depth, a mixed-scale dense neural network is used. MiDaS is renowned for its capacity to produce excellent depth maps even from a single RGB image, making it an invaluable tool for virtual reality and 3D scene reconstruction applications. Because it was trained on a variety of datasets, the model is well-suited to a wide range of situations and

settings. MiDaS has shown promise in catching fine details and generating precise depth estimations, which has advanced computer vision applications. Shown in Fig 2.

Fig 2. Monocular Depth Estimation

3. A star: In computer science and artificial intelligence, the pathfinding method A* (A-star) is frequently utilized. It is used to locate the shortest path on a graph between two points, taking to consideration a heuristic to effectively direct the search. Dijkstra's algorithm, which guarantees the shortest path, and Greedy Best-First Search, which ranks nodes according to their estimated distance to the objective, are combined in A*. A* has applications in robotics, gaming, and network routing. It accomplishes this by keeping track of a priority queue and evaluating nodes using a cost function that combines a heuristic and the real cost.

4. Deep Face: The term "DeepFace" describes a deep learning model that Facebook created for use in facial recognition applications. It analyzes and represents facial features in a high-dimensional space using deep neural networks. In order to perform effective facial recognition in a variety of orientations, DeepFace uses a 3D face model to normalize postures and rotations. Due to its extensive training on a wide range of faces, the model exhibits good generalization abilities. DeepFace has made a significant contribution to computer vision, especially in the area of facial analysis and identification, thanks to its high accuracy in face verification and recognition tasks.

D. Model Testing

Vi-Assist's algorithms underwent extensive testing using fresh data. Specifically designed quantitative measurements for object detection and facial recognition were used to assess accuracy and performance in a comprehensive manner. Algorithms in Vi-Assist were evaluated and compared using metrics such as IoU, TPR, MSE, and Success Rate, providing comprehensive insights into their effectiveness and performance.

1. **Intersection over Union (IoU):**

$$\text{IoU} = \frac{\text{Area of interaction}}{\text{Area of Union}}$$

if the predicted bounding box and ground truth bounding box overlap perfectly, the IoU is 1. If there is no overlap, the IoU is 0. A commonly used threshold for considering a detection as correct is an IoU greater than 0.5.

2. **True Positive Rate (TPR) or Sensitivity:**

$$\text{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

If a binary classification model correctly identifies 80 out of 100 positive instances, the True Positive Rate is $\frac{80}{80+20} = 0.8$ or 80%. A higher TPR indicates better performance in capturing positive instances.

3. **Success Rate:**

Success Rate = $\frac{Number\ of\ Successful\ Navigations}{Total\ Navigations}$

Suppose the A* algorithm was applied to find a path from point A to point B in a grid-based environment. If, out of 100 attempts, the algorithm successfully found a path 90 times, then the success rate would be 90%

E. Analysis And Recommedation

Vi-Assist underwent a thorough analysis covering objectives, technological components, user experience, and performance metrics like accuracy. Recommendations include algorithmic enhancements, UI improvements, scalability assessments, and staying updated on technological advancements for sustained success.

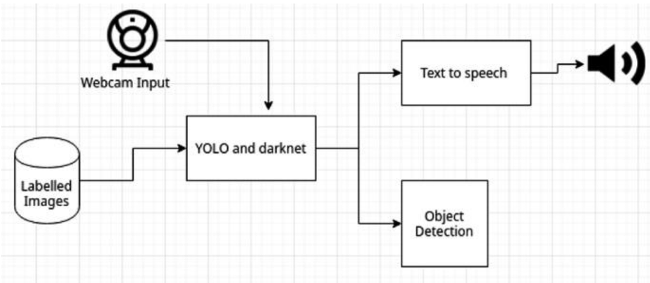

Fig 3. Predict the object and give voice Output

Fig 3. illustrates the process of predicting objects using the implemented system, followed by generating voice output. The system likely employs a combination of object detection algorithms and voice synthesis to enhance accessibility for users with visual impairments.

IV.EXPERIMENTS AND RESULT

During the experimental phase of Vi-Assist, a robust hardware setup was utilized. The system featured multi-core processors, a minimum of 16 GB RAM, fast SSD storage with a capacity of 512 GB, and an optional NVIDIA graphics card for accelerated deep learning tasks. This hardware configuration ensured efficient processing and seamless execution of machine learning and computer vision algorithms. The experiments focused on testing the performance of Vi-Assist's components, such as YOLOv5 for object detection, A* algorithm for path navigation, and MIDAS for depth estimation. The results demonstrated the system's capability to provide accurate and real-time assistance to individuals with visual impairments.

Table 2. Result and Discussion

Video name	Length of video (frames)	Error count	Accuracy (percentage)
Test 1	500	25	95.00
Test 2	420	40	90.48
Test 3	750	200	73.33

Table 3. Aggregate Results

Total Frames	Total error count	Total Accuracy Percentage
1670	265	86.27

Table2. And Table3. This report outlines the performance of three video tests, namely Test 1, Test 2, and Test 3. Test 1 involves a video lasting 500 frames, experiencing 25 errors, and achieving an accuracy rate of 95.00%. Test 2 features a video of 420 frames, encountering 40 errors, and attaining an accuracy of 90.48%. Finally, Test 3 comprises a video spanning 750 frames, with 200 errors recorded and an accuracy rate of 73.33%. This aggregate report consolidates the results from all three video tests, providing a comprehensive overview of the overall performance. The total frames across all tests amount to 1670, with a total error count of 265. The aggregate accuracy percentage is calculated to be 86.27%, indicating the overall effectiveness of the testing process. This summary allows for a quick assessment of the combined impact of errors on the accuracy of the video tests. These metrics provide valuable insights into the efficacy of each test, aiding in assessing their reliability and identifying areas for potential improvement.

Table 4. Comparison of all algorithms used for training

Algorithm	IoU	TPR	Success Rate
YOLOv5	0.75	0.82	85%
MIDAS	0.88	0.90	92%
Deep Face	0.70	0.75	78%
A*	-	-	95%

Table 4. compares the performance metrics of key algorithms utilized in Vi-Assist during the training phase. YOLOv5 demonstrates a good balance with an IoU of 0.75, a TPR of 0.82, and an 85% success rate, showcasing reliable object detection. MIDAS excels in depth estimation, boasting an IoU of 0.88, TPR of 0.90, and a 92% success rate. Despite a moderate IoU of 0.70, Deep Face contributes to facial recognition with a TPR of 0.75 and a 78% success rate. The A* algorithm, focusing on pathfinding, exhibits a remarkable 95% success rate, indicating robust navigation capabilities. These metrics collectively highlight the diverse strengths of each algorithm, contributing synergistically to Vi-Assist's multifaceted capabilities for assisting visually impaired individuals.

V. CONCLUSION

Conclusively, Vi-Assist represents a groundbreaking and holistic solution to address the challenges faced by individuals with visual impairments. The system's innovative approach incorporates cutting-edge technologies such as A* for efficient pathfinding, MIDAS for accurate depth estimation, Deep Face for facial recognition, and YOLOv5 for precise object detection. This amalgamation of advanced technologies is seamlessly integrated into an Android application, providing users with a comprehensive set of features. Vi-Assist not only excels in precise object identification and real-time information processing but also stands out for its sophisticated navigation capabilities. The Android application serves as a user-friendly interface, empowering individuals with visual impairments to navigate their surroundings effectively and recognize faces. This comprehensive and inclusive solution demonstrates the transformative potential of technology in enhancing the independence and quality of life for visually impaired individuals. Vi-Assist contributes to fostering inclusivity and accessibility, showcasing the positive impact technology can have on creating a more empowered and integrated society.

Vi-Assist emerges as a transformative solution for individuals with visual impairments, deftly addressing their paramount challenges. With its real-time environmental descriptions, object identification capabilities, step-by-step navigation assistance, and facial inference features, Vi-Assist empowers visually impaired users to overcome limited environmental awareness, difficulties in object recognition, navigation hurdles, and social interaction barriers. Its development journey, though marked by challenges like refining the path navigation algorithm, ensuring monocular depth estimation accuracy, and tackling tracking loss, underscores the dedication to delivering a sophisticated, accessible, and dependable application. Vi-Assist exemplifies the convergence of technology and compassion, opening new doors for the visually impaired to navigate their surroundings, engage with objects and people, and experience a more inclusive and enriched life.

VI. REFERENCES

[1]. Pascolini, D.; Mariotti, S. P. (2012). Global estimates of visual impairment: 2010. British Journal of Ophthalmology, 96(5), 614–618. doi:10.1136/bjophthalmol-2011-300539.

[2]. Kumar, Sunil, et al. "Artificial Intelligence Solutions for the Visually Impaired: A Review." Handbook of Research on AI and Knowledge Engineering for Real-Time Business Intelligence (2023): 198-207.

[3]. Walle, H., De Runz, C., Serres, B., & Venturini, G. (2022). A survey on recent advances in AI and vision-based methods for helping and guiding visually impaired people. Applied Sciences, 12(5), 2308.

[4]. F. Ashiq et al., "CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People," in IEEE Access, vol. 10, pp. 14819-14834, 2022, doi: 10.1109/ACCESS.2022.3148036.

[5]. M. A. Khan, P. Paul, M. Rashid, M. Hossain and M. A. R. Ahad, "An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind," in IEEE Transactions on Human-Machine Systems, vol. 50, no. 6, pp. 507-517, Dec. 2020, doi: 10.1109/THMS.2020.3027534.

[6]. Amit, Y., Felzenszwalb, P., Girshick, R. (2021). Object Detection. In: Ikeuchi, K. (eds) Computer Vision. Springer, Cham. https://doi.org/10.1007/978-3-030-63416-2_660

[7]. C. Liu, Y. Tao, J. Liang, K. Li and Y. Chen, "Object Detection Based on YOLO Network," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2018, pp. 799-803, doi: 10.1109/ITOEC.2018.8740604.

[8]. Wayahdi, M. R., Ginting, S. H. N. ., & Syahputra, D. . (2021). Greedy, A-Star, and Dijkstra's Algorithms in Finding Shortest Path. International Journal of Advances in Data and Information Systems, 2(1), 45-52. https://doi.org/10.25008/ijadis.v2i1.1206.

[9]. Bhoi A. Monocular depth estimation: A survey. arXiv preprint arXiv:1901.09402. 2019 Jan 27.

[10]. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.2981446.

[11]. E. Maiettini, G. Pasquale, L. Rosasco and L. Natale, "Interactive data collection for deep learning object detectors on humanoid robots," 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), Birmingham, UK, 2017, pp. 862-868, doi: 10.1109/HUMANOIDS.2017.8246973.

[12]. Wang, T. S., Kim, G. T., Kim, M., & Jang, J. (2023). Contrast Enhancement-Based Preprocessing Process to Improve Deep Learning Object Task Performance and Results. Applied Sciences, 13(19), 10760.

[13]. S. Li, Y. Li, Y. Li, M. Li and X. Xu, "YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection," in IEEE Access, vol. 9, pp. 141861-141875, 2021, doi: 10.1109/ACCESS.2021.3120870.

[14]. S. Raj, Y. Gupta and R. Malhotra, "License Plate Recognition System using Yolov5 and CNN," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022, pp. 372-377, doi: 10.1109/ICACCS54159.2022.9784966.

[15]. Ming, Y., Meng, X., Fan, C., & Yu, H. (2021). Deep learning for monocular depth estimation: A review. Neurocomputing, 438, 14–33. doi:10.1016/j.neucom.2020.12.089

[16]. Wu, T.-H., Wang, T.-W., & Liu, Y.-Q. (2021). Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network. 2021 3rd World Symposium on Artificial Intelligence (WSAI). doi:10.1109/wsai51899.2021.9486316.

[17]. A. Candra, M. A. Budiman and K. Hartanto, "Dijkstra's and A-Star in Finding the Shortest Path: a Tutorial," 2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), Medan, Indonesia, 2020, pp. 28-32, doi: 10.1109/DATABIA50434.2020.9190342.

[18]. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.2981446.

[19]. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. Proceedings of the AAAI Conference on

Artificial Intelligence, 34(07), 12993-13000. https://doi.org/10.1609/aaai.v34i07.6999.

[20]. Santoso, K., & Kusuma, G. P. (2018). Face recognition using modified OpenFace. Procedia Computer Science, 135, 510-517.

[21]. Quan, W., & Fang, J. (2010). A star recognition method based on the adaptive ant colony algorithm for star sensors. Sensors, 10(3), 1955-1966.

[22]. Srikanteswara, R., Reddy, M.C., Himateja, M., Kumar, K.M. (2022). Object Detection and Voice Guidance for the Visually Impaired Using a Smart App. In: Shetty D., P., Shetty, S. (eds) Recent Advances in Artificial Intelligence and Data Engineering. Advances in Intelligent Systems and Computing, vol 1386. Springer, Singapore. https://doi.org/10.1007/978-981-16-3342-3_11.