2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)

Hybrid Image Processing Device as Wearable Aide for Visually Impaired

Ashish Papanai
Department of Computer Science and Engineering
Maharaja Agrasen Institute of Technology
Delhi, India
ashishpapanai00@gmail.com

Harsh Kaushik
Department of Electrical and Electronics Engineering
Maharaja Agrasen Institute of Technology
Delhi, India
harshkaushik717@gmail.com

***Abstract*— With the advent of technology and algorithms in image processing, many wearable aides are available in the market, and researchers across the globe are developing new solutions. The existing solutions and products fail to provide an economically viable and single solution to the problem of visually impaired people. This study introduces AIde, a novel system amalgamating optimized computer vision algorithms and wearable IoT devices to convert the information extracted from the camera mounted on the user to audio-based signals for assisted mobility. The proposed system has three user modes: indoor, outdoor, and hybrid, which can be modified using the buttons provided on the wearable device. The indoor mode can be used for object detection and classification to increase the accessibility of the objects inside a confined space, and the outdoor mode will provide the user with enhanced mobility and directional sense based on the highlighted path captured by the mounted camera. The hybrid model combines both the features of indoor and outdoor modes, which will detect objects and improve mobility simultaneously. The proposed model is optimized to work in a computationally efficient environment with an at par efficiency and accuracy as the state-of-art and costly market alternatives.**

Keywords—Computer Vision, Parallel Computing, Video Processing, Edge Computing, Wearable Devices, Internet of Things, Image Processing

I. INTRODUCTION

The data of 2020 from the International Agency for the Prevention of Blindness ranks India at the top of the list of visually impaired people [1]. The average income of an Indian household, as reported by moneymint.com, is Rs 16,000 (220 USD) [2], and this income reduces significantly in-case of differentially abled or impaired individuals. Moving around the world, China and India account for more than a billion visually impaired people. The economically weaker sections of Africa account for more than 32 million visually impaired people [3]. As shown in Fig 1, Most visually impaired people live in developing nations with nominal incomes and cannot afford a guide or resources to invest in costly wearable aides available in the market. The most inexpensive device currently available for the visually impaired costs around 3000 USD, including features like text reading, face recognition, and other basic features. The existing devices ignore the most crucial aspect of eyes and vision: mobility. The other commonly available market solutions which focus on mobility fail to address the concept of vision and joy of explaining what the eyes can see. Various hand-held aides are available, ranging from 1000USD to as high as 10000USD, which is far too expensive to be considered in an average household of the developing countries.

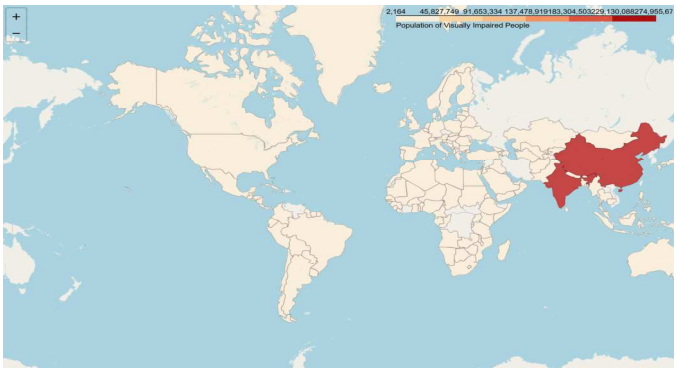

Fig. 1. Geographical distribution of the visually impaired people

This paper addresses the ineffectiveness of the expensive market solutions and presents a novel, hybrid, and optimized Artificial Intelligence Guide (AIde) for visually impaired people. The prototype presented in this model aims to solve three major problems of wearable-camera aides: cost, mobility, and privacy.

The advent of cloud-based deep learning strategies [4] poses a severe threat to individuals' security and privacy, streamed continuously to and from the cloud-based data analyzers. This study proposes an end-to-end model that provides real-time, fast, in-place analysis of the captured user frames. To make the system more robust and cost-effective, purely mathematical and statistical models for lane detection and curvature analysis are used, which takes less than 10 seconds to determine the curvature of the lane based on the sequence of frames obtained from the camera module. The Object Detection module for the indoor mode of AIde requires minimal GPU capabilities as it uses a Convolution Neural Network for object detection and image captioning.

The results generated from this model are comparable with the state-of-art and the available market products but with a significant reduction in computational requirements, time, and costs.

This paper is divided into six sections; Section II focuses on the previous work and its limitations. Section III is a detailed analysis of the data, use-cases, and background. Section IV describes the system architecture and the IoT device. Section V focuses on the results obtained from the proposed model and

compares various object detection techniques. Section VI concludes the paper and provides future scope for the study and findings.

II. Related Work and Limitations

In section I of the paper, we provided an overview of the previous work done in Computer Vision for the visually impaired. Our work connects several pieces of literature and creates an optimized blend of the existing solutions. The proposed solution is intended to provide an economical option for visually impaired people.

MAVI by Kedia et al. [4], a breakthrough computer vision solution for the visually impaired, provides clean annotated data for future studies, but their solution fails to discuss the user's privacy concerns using their wearables. They also fail to provide a solution for regions with little to no internet connectivity. The cloud-based solutions are not reliable, considering that even car manufacturers (who deal with mobility) are moving to non-internet-based geopositioning systems rather than internet-based path locations.

The second major problem of models dealing only in pedestrian lane detection, the existing models focus on extracting the patches of interest (POI) ignoring the mathematical computations to predict the curvature of the highlighted path, the RANSAC technique introduced by Le et al. [5] doesn't provide a solid framework for using the data obtained from the highlighted path. Similarly, various other lane and object detection solutions provide a simple solution by highlighting/ detecting the already highlighted path, thereby failing to provide a broader solution to the problem of lane detection and curvature prediction in the absence of highlighted path [5], [6].

For this study, we studied the results from Tan et al. [7], the study by RetinaNet, and various versions and decided to optimize the versions further to reduce the computational cost. Even though RetinaNet provides better mean average precision as compared to YOLO v3 and Single Shot Detection (SSD), YOLOv3 works better in the prediction of multiple objects by predicting multiple bounding boxes and their categories, the detection speed is much faster, and computational cost is minimized as compared to other model architectures. And recent models of YOLO, that is, YOLO v4, v5, PP-YOLO, provide a mean average precision of 83.5% and three times faster results thanetinaNet [7]–[9].

Comparison of RetinaNet for indoor object classification by Afif et al. [10] proves the mean average precision of the model as 84.61%, which is comparable with our model architecture, but it fails to match the speed and computational requirements. The model also fails to provide a single-shot solution for the problem of providing an AI guide for visually impaired people.

GOLD by Bertozzi et al. [11] provides a similar solution to AIde introduced in this study, but it fails to provide efficient and reliable audio-based outputs based on the scenes captured by the mounted camera. It focuses more on obstacles and lane detection to reduce accidents while driving. We have successfully pivoted the idea to provide the visually impaired with reduced mobility to sit and drive their wheelchair or any other mobility aid [11].

The main contribution of our work is to provide the community with a scalable, efficient, and inexpensive image and video processing framework. The following features are highlighted in the study:

1. Comparison of various object detection architectures, based on accuracy, computational requirements, and computational time
2. Providing a detailed analysis of the system architecture of AIde.
3. Implement a simple prototype of the proposed system with a brief outline and functions of the used components.

III. Algorithm And Data

A. Problem Formulation

This study aims to develop an optimized object detection and captioning model for indoor object detection. The generated captions are written in a text file, from where the pyttsx3 module converts the text output to audio signals. These audio signals will be played in the user's headphones to improve the indoor mobility of the visually impaired user.

For enhanced outdoor mobility, real-time video is provided to the lane detection and curvature prediction module for enhanced mobility in an outdoor environment.

The third mode of AIde is a hybrid indoor-outdoor mode, which amalgamates the function of lane detection and objects captioning. This module is intended for the user to move safely and experience the environment through their virtual guide.

We aim to provide a single-shot solution to the community and an economically viable device for visually impaired people through this solution. This device will work with the same accuracy and prediction capabilities in indoor and outdoor environments.

B. Algorithm

1) ***Outdoor (Mobility) Mode:*** We start with color selection for outdoor lane detection based on the color of interest. This color depends on the color of the highlighted path for visually impaired people.

Fig. 2. Extraction of meaningful features from the captured image (a) original image (b) canny's object detection (c) captured ROI

Fig 2(b) shows the extracted information about the coordinates of the highlighted path for the visually impaired. The regression model later uses this information to predict the curvature of the path for enhanced mobility.

After selecting the highlighted path, the module finds the region of interest to run the statistical model for curvature prediction. The captured region of interest for the study is shown in Fig 2(c). Finally, we use Hough Transform on the image with the edges detected by Canny's Algorithm.

A line can be represented with the following equation:

$$y = a.x + b \quad (1)$$

The form represented in equation (1) fails to work for vertical lines, so we use the following format:

$$r = x.\cos\theta + y.\sin\theta \quad (2)$$

The lane detection module performs the following steps:

1. Edge detection, using Canny edge detector
2. Selection of Region of Interest (ROI).
3. Mapping of edge points to the Hough space and storage in an accumulator.
4. Interpretation of the accumulator to yield lines of infinite length.
5. Conversion of infinite lines to finite lines.
6. Calculating the point of intersection of the two finite lines corresponding to the path.
7. If the point of intersection (POI) is in the center of ROI, the user is told to go straight. If the POI is left or right, the user is told to go to left or right, and vice versa.

*2) **Indoor (Object Detection) Mode:*** The indoor edge detection module uses a much simpler transfer learning approach from YOLO v5. The final/output layer of the model is replaced according to the 75 classes considered in the study. The model is optimized and improved using the methods suggested by Zhu et al. [12]. The method suggested for optimized drone captured scenarios is transfer learned for indoor Object captioning by AIde.

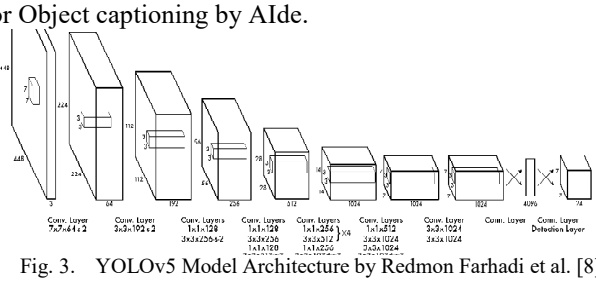

Fig. 3. YOLOv5 Model Architecture by Redmon Farhadi et al. [8]

The model is trained on the COCO dataset and provides us with a MAP of 84.9% when tested with the dataset of MAVI by Kedia et al. [4]. Before deciding to go forward with YOLO v5 for object detection and captioning, we tested various versions of YOLO and other object detection models like ResNet, SSD, et Cetra. Fig 4 shows the comparison plots of various YOLO v5 versions and EfficientDET.

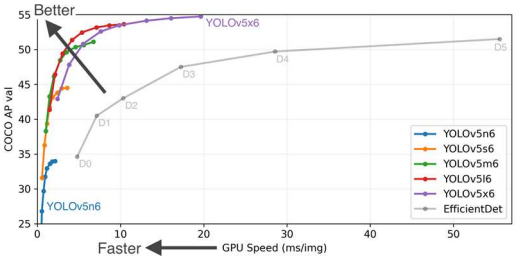

Fig. 4. Comparison plots of YOLOv5 (various versions) and EfficientDET, the results are replicated from the GitHub repository of YOLOv5 [13].

*3) **Hybrid Mode:*** AIde's hybrid mode (Outdoor + Indoor) uses thread-based parallelism provided in python, and three threads parallelly execute the Object captioning, Lane detection, and the text to speech module for efficient working of the developed system. The thread execution is accelerated by using PyCUDA for parallel computation.

C. Data

For making the model prepared for unseen data in a real-time environment, we have used a blend of data obtained from various sources, and at the same time, we have worked on our personalized AIde dataset as well. The object detection module of AIde is trained using Common Object in Context (COCO) database [14]. For testing the accuracy of the model and its suitability in the Indian market, we have tested it using the MAVI dataset provided by Kedia et al. [4]. Our homegrown video database used for the Outdoor Mode (Lane Detection) is based on 500 real-time dashboard camera videos. The model is tested using fifty validation videos recorded from various streets, malls, and metro stations of New Delhi with the highlighted path for the visually impaired.

Table 1 shows the description of images per classes of interest obtained from the COCO dataset and the images obtained from the MAVI Dataset for training and testing AIDE.

Fig. 5. Images from the COCO Dataset (a) apple (b) person (c) food (d) Sign Board (Stop)

TABLE I. Description Of COCO Dataset (Training and Test)

S. No.	*Class Label*	*Number of Images*
1	Person	66,808
2	Bicycle	3,401
3	Car	12,786
…	….	…
70	Pizza	3,319

[a.] Only classes of particular interest

The total 123,287 images are split into training (80%) data, and testing (20%) data are obtained from COCO Dataset. Table II. shows the description of the images obtained from the MAVI Dataset.

Fig. 6. Images from the MAVI Dataset (a) sign-board (b) dog (c),(d) cow

TABLE II. Description Of MAVI Dataset (Validation)

S. No.	*Class Label*	*Number of Images*
1	Dog	1,498
2	Cow	1,604
3	Sign Board	1,493

The images obtained from the MAVI Dataset validate the proposed product for Indian markets. Table III. Shows the distribution of the AIde video dataset for testing the mathematical lane detection algorithm.

Fig. 7. Frames extracted from the videos of AIde Dataset

TABLE III. Description Of AIde Video Dataset

S. No.	*Video Mode*	*Number of Videos*
1	Day	400
2	Night	100
3	Validation	50

The AIde dataset comprises 20-80% testing and training data. Fifty validation videos are used to check the accuracy of the product for the Indian market and users.

The AIde dataset comprises a variety of texture-based videos of highlighted paths for the visually impaired, which trains the model for a robust application in real-world scenarios. Our outdoor module works on a principle similar to that of a self-driving AI, and we have trained and tested the mathematical and statistical approach on reading time driving videos. The data is pre-processed based on the algorithm explained in Sub Section A of Section III.

IV. AIde: System Architecture

AIde is powered by NVIDIA Jetson Nano, which provides a power-efficient computer to run multiple neural networks and machine learning modules in parallel. Real-time video is recorded from the wearable camera mounted on the user. AIde uses a 4K (Ultra High Definition) Alluvium camera kit provided by Allied Vision. The real-time video recording from the video is stored in the SD Card (128 GB) and is cleared after 60 seconds by a cron-job. The SD Card (Memory module) also comprises NVIDIA JetPack SDK, including the OS images from NVIDIA Jetson, AIde uses Ubuntu 20.04LTS, and Linux Kernel 5.4. NVIDIA Drivers for GPU and Allied Camera Module are included in the system image. Audio signals are retrieved from the Jetson module using USB Headphone. Wired headphones reduce the chances of delay in audio which is possible in Bluetooth headphones. The battery module of AIde comprises a 25000 mAh Krisdonia Portable Laptop Charger, which NVIDIA Jetson supports. The power module has sufficient power, can provide continuous usage up to 48 hours, and has an LED display to check the battery percentage.

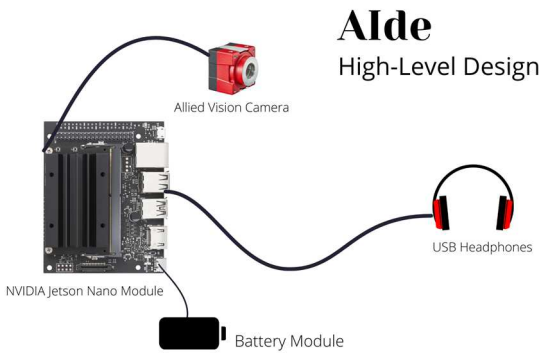

Fig. 8. High-Level Design of AIde

Fig 8 shows the high-level architecture of AIde, the costing of the components is shown in Table IV.

TABLE IV. Costing Of Components Used In AIde

S. No.	*Component*	*Cost (USD)*
1	NVIDIA Jetson Nano	59
2	Alluvium Allied Vision Camera	229
3	USB Headphones	67
4	Battery Module	95
5	Structure (And miscellaneous)	50

The total cost of AIde components is 500 USD, which can reduce significantly in the case of bulk production.

AIde works by capturing the video from the camera module, which is transferred to the processing module (NVIDIA Jetson Nano). The processing module runs the algorithm described in Section III based on the mode selected by the user. The results generated by the algorithm and the object detection module are written into the text file, then read

by the text-to-speech library and sent to the user through headphones (audio module). Figure 9 Shows the level-2 architecture of AIde.

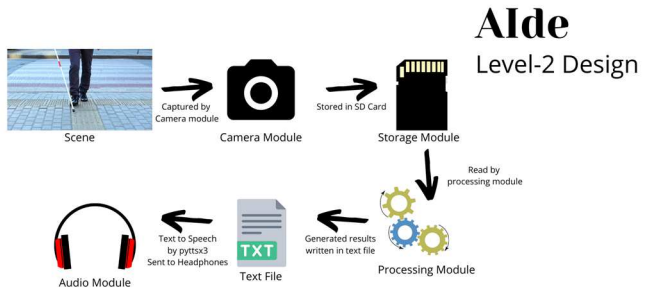

Fig. 9. AIde Level-2 architecture (Pictorial representation of steps followed)

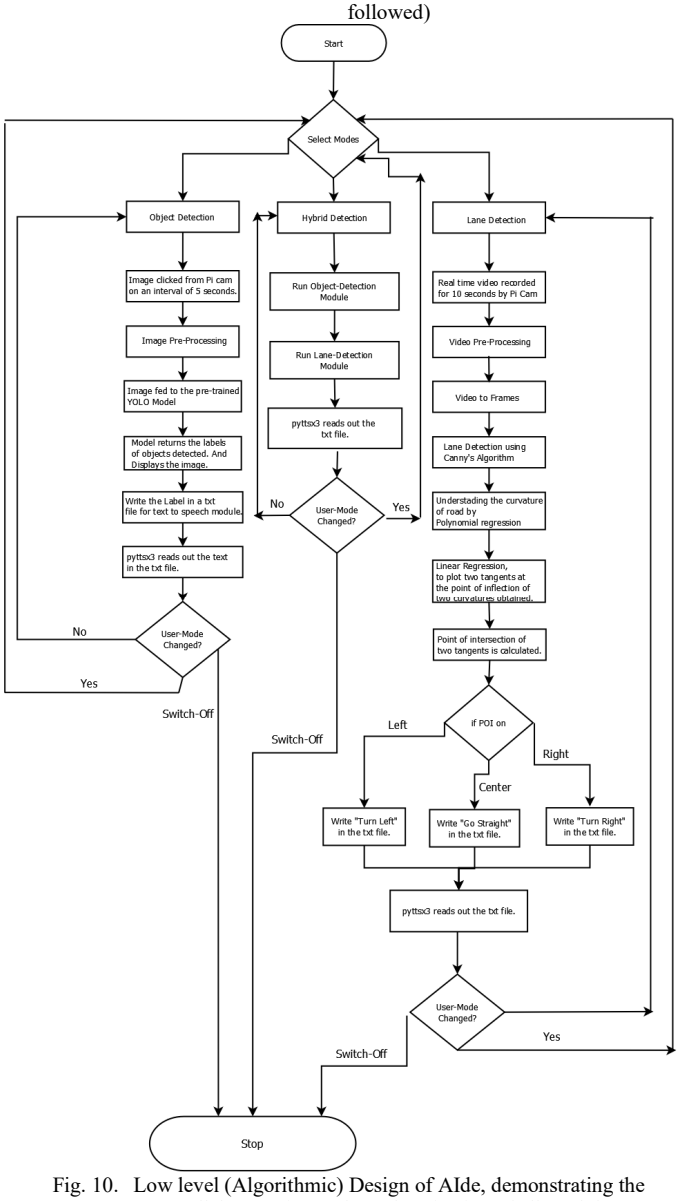

Fig. 10. Low level (Algorithmic) Design of AIde, demonstrating the parallel execution and synchronization of all modules

AIde runs locally and end-to-end in the wearable devices used by the user. This end-to-end model reduces any possibility of a data leak, which is possible in the cloud implementation of wearable devices. The data collected from the real-time camera module is stored only for 60 seconds in the memory module. This reduces the user's chances of any security threat as his real-time positioning will be available only to him and his guide (AIde).

Figure 10. shows the low-level design of the AIde algorithm, highlighting how we have incorporated threading to implement parallel execution of the processing, audio, image captioning, and lane detection and prediction module. The system architecture explains how this study has successfully provided a cost-effective and optimized computer vision solution to enhance the mobility of visually impaired people.

V. Results and Evaluations

This section of the study presents the accuracy, time, and energy requirements of the solution described. Table V of this section compares various sub-versions of object detection model architecture (YOLOv5) tested for implementation of AIde.

TABLE V. Comparison YOLOv5 Sub-Versions for AIde

Model	*Size (pixels)*	*mAP*	*Speed (ms)*
YOLOv5n	640	46.0	6.3
YOLOv5n6	1280	50.7	8.1
YOLOv5m6	1280	69.0	11.1
YOLOv5l6	1280	72.3	26.2

The accuracy of the lane detection and prediction module, which is supported by the Hough Transform algorithm, is 84.25% accurate based on the sub-pixel location.

Table VI shows the precision and recall obtained from the YOLOv5 model trained on COCO, MAVI, and AIde dataset.

TABLE VI. Evaluation Metrics for AIde Indoor Module

Metrics	*Results*
mAP (0-0.5)	0.79
mAP (0.5-0.95)	0.45
Precision	0.85
Recall	0.79

Figure 11 shows the results generated from the lane detection module for autonomous driving of the assisted wheelchair and lane detection of the highlighted path.

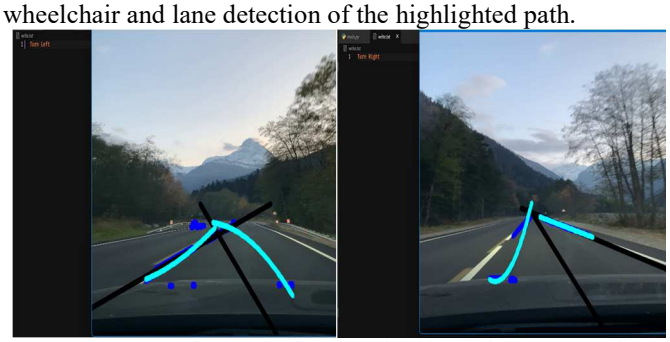

Fig. 11. (a), (b) Outputs from lane detection and curvature prediction module

The left-hand side of the video shows the text file, which will be converted to speech by the pyttsx3 module. Figure 12 (a) (b) shows the outputs generated from the indoor mode (object detection module).

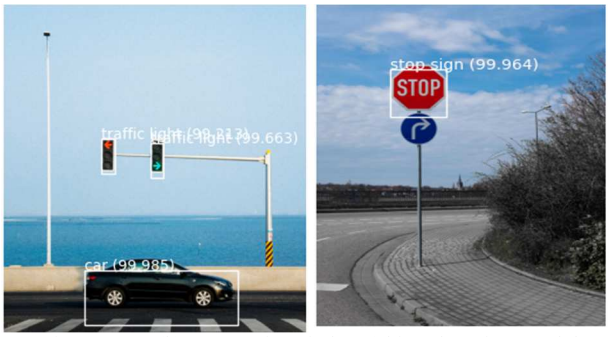

Fig. 12. (a), (b) Outputs from indoor (object detection) module, used in a hybrid scenario

AIde uses power as low as 5W, and the 25000 mAh battery module provided is expected to last for 48 hours (continuous usage) in a single charge. We have significantly reduced the data transfer rates by using an efficient microSD card (storage module). The solution is latency-free as we don't require uploading or downloading data to any cloud-based servers. The 128 Core NVIDIA Maxwell architecture GPU is sufficient to efficiently run the object detection neural network, evident from the results presented in Table V.

VI. Conclusion and Future Work

AIde has significantly reduced the computational complexity, time, and energy requirements and thereby provides the community and users an efficient solution for the mobility and vision of the visually impaired. Compared to the entry-level options available in the market, AIde offers a wide variety of features and modes at a much reasonable and reduced financial cost. The hardware used in AIde is the best available option in the market, considering the 4K camera for video capture and the NVIDIA Jetson Nano for running deep neural networks.

As a part of future work, we aim to add features like reading assistance for the visually impaired and provide a haptic response to the user based on the texture of the surface of the road/ place where the user is walking. These features are intended to take us closer to the goal of enhanced mobility of the visually impaired using Artificial Intelligence and Image Processing.

References

[1] "Country Map & Estimates of Vision Loss - The International Agency for the Prevention of Blindness." https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/countries/ (accessed Feb. 06, 2022).

[2] "What is Average Salary in India? (2022) - MoneyMint." https://moneymint.com/what-is-average-salary-in-india/ (accessed Feb. 06, 2022).

[3] Silvio P. Mariot and World Health Organisation, *GLOBAL DATA ON VISUAL IMPAIRMENTS 2010*. 2010. Accessed: Feb. 06, 2022. [Online]. Available: https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf

[4] R. Kedia *et al.*, "Mavi: Mobility assistant for visually impaired with optional use of local and cloud resources," *Proceedings - 32nd International Conference on VLSI Design, VLSID 2019 - Held concurrently with 18th International Conference on Embedded Systems, ES 2019*, pp. 227–232, May 2019, doi: 10.1109/VLSID.2019.00058.

[5] M. C. Le, S. L. Phung, and A. Bouzerdoum, "Pedestrian lane detection for the visually impaired," *2012 International Conference on Digital Image Computing Techniques and Applications, DICTA 2012*, 2012, doi: 10.1109/DICTA.2012.6411701.

[6] T. Supriyadi, B. Setiadi, and H. Nugroho, "Pedestrian lane and obstacle detection for blind people," *Journal of Physics: Conference Series*, vol. 1450, no. 1, p. 012036, Feb. 2020, doi: 10.1088/1742-6596/1450/1/012036.

[7] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–11, Dec. 2021, doi: 10.1186/S12911-021-01691-8/TABLES/4.

[8] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." [Online]. Available: https://pjreddie.com/yolo/.

[9] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, pp. 687–694, Mar. 2020, doi: 10.1109/ICACCS48705.2020.9074315.

[10] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, and M. Atri, "An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation," *Neural Processing Letters 2020 51:3*, vol. 51, no. 3, pp. 2265–2279, Jan. 2020, doi: 10.1007/S11063-020-10197-9.

[11] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998, doi: 10.1109/83.650851.

[12] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2021-October, pp. 2778–2788, Aug. 2021, doi: 10.1109/ICCVW54120.2021.00312.

[13] G. Jocher *et al.*, "ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support," Oct. 2021, doi: 10.5281/ZENODO.5563715.

[14] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 740–755, May 2014, Accessed: Oct. 12, 2021. [Online]. Available: https://arxiv.org/abs/1405.0312v3