ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

Design Research on the Application of YOLOv5-based Road Condition Detection in Assistive Devices for the Visually Impaired

Zihan Ma

School of Educational Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China

Abstract: *Outdoor navigation has long been a wicked problem for the visually impaired. With the development of technologies, new designs with them are temping to deal with existing obstacles both from within and outside, but still leaving pain points to be solved. Driven by social considerations and user-centred perspectives, we propose an innovative and sustainable design framework for assistive devices catering to the visually impaired in this project. The core implementation layer utilizes Arduino, while the decision-making layer employs the YOLOv5 algorithm for object detection tasks, enabling hardware interaction in response to the detection results. The device successfully provides features such as environmental information feedback, tactile paving search and privacy intrusion warning, transforming necessary visual information into auditory and tactile feedback. By controlling it at a lower cost and simplifying interaction methods, our device offers essential mobility assistance to a broader range of visually impaired users.*

Keywords: *Visually Impaired Assistive Device, Objection Detection, YOLOv5 Algorithm, Arduino, Human-Centred*

1. Introduction

The International Agency has projected for the Prevention of Blindness (IAPB) that the number of individuals worldwide with moderate to severe visual impairments will exceed 500 million by 2050. Focusing on the health and lives of visually impaired individuals has become an essential agenda for global sustainable development. Attempting to address the mobility challenges faced by visually impaired individuals serves as an entry point for connecting technology and daily life. It thus holds practical significance for exploring the application of road condition detection technology in assistive devices for the visually impaired.

With the advancement of technology, convolutional neural networks (CNN)[1] based image recognition algorithms and object detection algorithms have become essential fields in computer vision, making more direct contributions to human life and being applied in various domains. Among them, the YOLO (You Only Look Once)[2] algorithm is one of the recently emerged object detection algorithms. Benefiting from the trend of open design, numerous researchers have contributed to its development and made continuous progress on the foundation laid by others. Therefore, this study aims to explore the potential application of road condition detection technology based on YOLOv5 in assistive devices for the visually impaired. Furthermore, the study proposes a directly feasible design framework using design thinking. Integrating technological development and actual human needs will directly promote social sustainability.

2. Literature Research

Based on literature research, assistive designs for visually impaired individuals can be broadly categorized into (1) tactile-based systems, (2) guiding and environmental design, and (3) electronic devices.

Tactile-based systems refer to designs that convey information through touch and tactile feedback, including common designs such as tactile paving and tactile maps. Devices based on tactile feedback often rely on algorithms to convert environmental information captured by cameras into braille and then convert the corresponding information into binary form to be transmitted to related braille generation

ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

devices, presenting tactile feedback through mechanical or thermohydraulic[3] means.

Electronic devices encompass portable devices that utilize various sensors to detect obstacles, perceive the environment, and provide user feedback. These devices often rely on ultrasonic sensors or infrared sensors to quickly determine the user's spatial relationship with objects in the surroundings. When it comes to environmental designs for the visually impaired, apart from urban planning initiatives, there are also efforts focused on community redesign and the deployment of radio-frequency identification (RFID) technology[4] to achieve barrier-free mobility. While methods like this have the potential to address the problem fundamentally, they require significant resources and may have slow results, serving as a long-term vision for future smart living. The design of electronic devices holds the potential for making more direct contributions to the visually impaired in various aspects. Literature reading shows that many existing devices for the visually impaired can be broadly classified into wearable and mobile-guiding categories. Standard wearable devices include intelligent glasses, backpacks, and canes. The smart cane developed by Du Yushong et al.[5] As such, mobile-guiding designs often simulate the working patterns of guide dogs, such as the CaBot active guidance and obstacle avoidance robot developed by João Guerreiro's team at Carnegie Mellon University[6].

Most of these designs aim to address three major aspects of challenges visually impaired individuals face: environmental perception, obstacle avoidance, and navigation.

Currently, several mainstream object detection algorithms, such as Region-based Convolutional Neural Network (R-CNN), Single Shot MultiBox Detector (SSD), and YOLO algorithm, have been developed to tackle the problem of object localization and recognition using different approaches. However, they differ regarding evaluation indicators, such as accuracy and inference time. The YOLO algorithm stands out as lightweight, efficient, and accurate, which makes it well-suited for addressing the travel challenges of visually impaired individuals. In the ongoing development of the YOLO algorithm, YOLOv5 has gained more theoretical research foundations and experimental data, providing a more realistic and evidence-based approach and tools.

3. Design Theory and Conclusion

3.1 Design Thinking

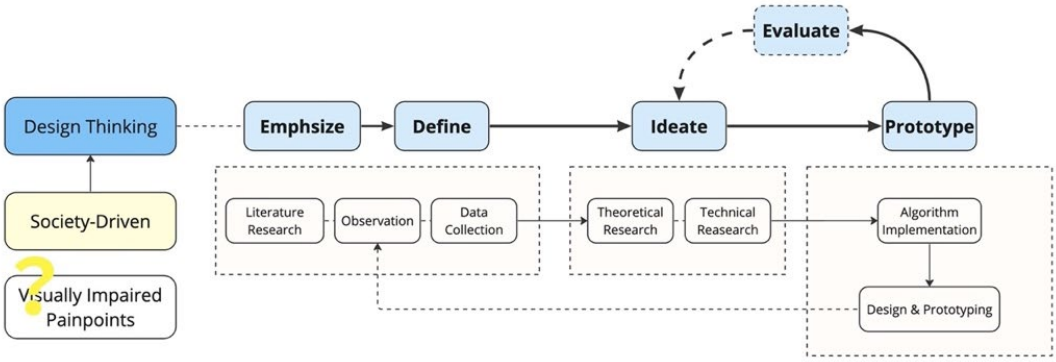

Figure 1: Overall Project Process

Richard Buchanan has proposed that design should be regarded as a contemporary liberal art [7]. Design thinking is the link that connects design to the real world. It can represent the designer's cognitive and strategic design processes and embody a designer's habitual design concepts or familiar design knowledge. According to Tim Brown from IDEO, who defined it as "a human-centred approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success[8]", it can be seen that design thinking is of significant importance to the design concept of this study. According to this, the overall project process is shown in Figure 1.

3.2 Insights

Based on information search and observations, it was found that visually impaired individuals often efficiently utilize their auditory abilities to make preliminary judgments about their surroundings. Common traditional mobility aids for the visually impaired include white canes and guide dogs. White canes are the most widely used and have a globally standardized appearance and usage. Meanwhile, the

ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

training cost for guide dogs is high, as presented by the International Guide Dog Federation (IGDF), in 2022, there are approximately 20,000 active guide dogs globally, with nearly 70% of them being used in major cities.

Furthermore, during on-site investigations conducted around the project's implementation area, it was discovered that within just a few kilometres, there were three different types of tactile paving blocks. After further consultation of relevant documents, it was found that an efficient and time-tested accessible design, documented in the standards for urban accessibility, provided a foundation for later optimization. However, due to inconsistent implementation of accessibility initiatives, it has posed challenges for design advancement.

From these design insights, it can be concluded that the project successfully identified real pain points through design practice, including (1) limitations of traditional mobility aids, (2) high cost and lack of sustainability of existing optimized intelligent devices, and (3) uncertainty in the effectiveness of urban planning execution. Overall, the project aims to provide more effective and accessible solutions for visually impaired individuals in their daily mobility, focusing on these identified pain points.

4. Implementation of YOLOv5 Algorithm

Based on the theoretical foundation discussed earlier, the YOLOv5 algorithm has been selected as the object detection algorithm for this project. The experiments will be conducted on a self-made dataset, and YOLOv5s and YOLOv5m will then be trained and tested separately. The results will be summarized and analyzed to obtain a practical algorithm training plan and trained models suitable for real-world scenarios. The algorithm implementation process for this chapter is illustrated in Figure 2.

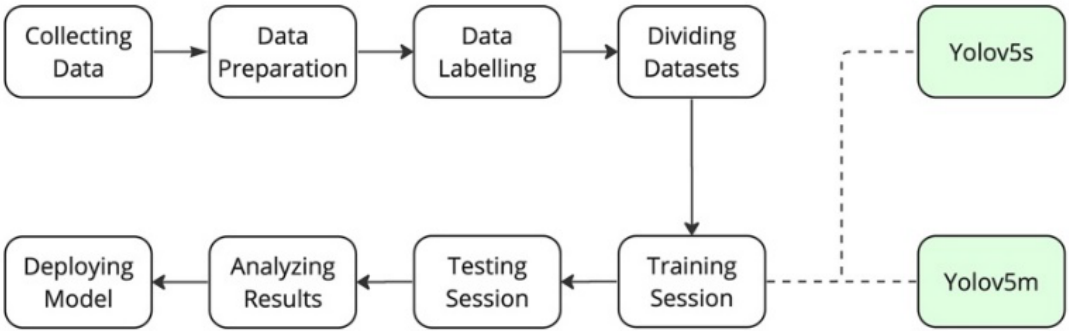

Figure 2: Algorithm Implementation Process

4.1 Data Preparation

Due to the project's focus on the daily mobility, the project created a self-made dataset called "Visona", which includes categories such as tactile paving, zebra crossing, and steps. The image data was collected from three methods, including existing relevant datasets, from the internet, and captured by phone camera in person. These images were then labelled using the Roboflow platform. The final self-made dataset "Visona" contains a total of 8,732 images. To ensure that the final results can directly meet the practical needs, the recognition categories should include key target classes for indoor and outdoor activities. Therefore, this project uses both the PASCAL VOC and Visiona datasets. The final dataset composition is presented in Table 1. The total dataset consists of 30,235 samples, divided approximately in the ratio of 5:4:1 for training, validation, and testing, as shown in Table 2.

Table 1: Final Dataset Composition

Datasets	Subsets	Pictures	Count	Total
Visona	train	7308	8732	**30235**
	valid	815		
	test	710		
PASCAL VOC	2007trainval	5011	21503	
	2007test	4952		
	2012trainval	11540		

ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

Table 2: Dataset Ratios

Subset	Datasets	Pics	Percentage
train	2007trainval	23859	0.51
	2012trainval		
	visona_train		
valid	2007trainval	17366	0.37
	2012trainval		
	visona_valid		
test	2007test	5662	0.12
	visona_test		

4.2 Evaluation Indicators

In object detection, performance will be calculated regarding evaluation indicators from multiple aspects. Precision is one of the critical metrics to evaluate the accuracy of the model's detections. It is relevant when the model's correct predictions are crucial for user safety during outdoor activities. However, increasing precision may lead to more false negatives, thus reducing recall and vice versa. These metrics evaluate the model's ability to detect true positives while minimizing false positives and false negatives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

Inference Time, also known as the model's prediction time, measures the average time it takes for the model to complete object detection inference on input images. Considering the practical application of this assistive device, it is likely to be a mobile device with limited hardware capabilities. Therefore, the model's complexity and processing time on devices without GPU should be considered, especially to meet the real-time processing requirement at 100 ms per frame. Additionally, exploring other techniques to compensate for the lack of GPU on local mobile devices may be necessary.

In summary, for this project's consideration of the future use of the algorithm in assistive devices for the visually impaired, preliminary model training and testing should focus on achieving the ability to detect objects in real-world scenarios. Thus, the expected results are presented in Table 3.

Table 3: Performance Requirements

Performance Measures	Requirement
Precision	≥80%
Inference time	≤100ms

4.3 Training Session

Comparative experiments were conducted in different configurations during the model training phase to explore whether the model can be effectively applied to assistive devices for the visually impaired. The device configurations provided by Kaggle and the detailed information on the local machine are shown in Table 4. The performance of the models was recorded, and ultimately the same complexity-level models were trained on both Machine A and Machine C.

Table 4: Machines' Configurations

Machine	Type	Accelerator	VRAM
A	None	CPU	-
B	GPU T4 x2	Tesla T4	15110MiB
C	GPU P100	Tesla P100-PCIE-16GB	16281MiB
D	None	Intel UHD Graphics 630	1536

The model training process used the dataset from the training set described in the previous section. The optimizer used was SGD, and the image size was 640. The batch size was set to -1, meaning AutoBatch was used to adaptively adjust the batch size based on the GPU's memory capacity and system performance. The initial iteration count was set to 200, and the hyperparameters used default values. Pre-trained models were chosen to meet real-world usage requirements. The training duration and the average accuracy on the validation set for both training sessions are shown in Table 5.

According to the results, YOLOv5m requires longer training time and incurs higher implementation costs due to its more complex and resource-intensive nature. However, the trade-off is that YOLOv5m can achieve higher accuracy. Considering the model's performance and training resource consumption, it

ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

is essential to prioritize ensuring that the model can provide reliable target detection functionality for the visually impaired in real-world scenarios.

Table 5: Training Results

Model	Approximate Time	Average Precision (Validation)	Average Precision (Visona-Validation)
YOLOv5s	**72h**	**0.80**	**0.85**
YOLOv5m	96h	0.83	0.86

4.4 Testing Session

After training, we obtained the best and final iteration weights for both YOLOv5s and YOLOv5m models. The best weights are determined by computing a fitness function w = [0.0, 0.0, 0.1, 0.9], which corresponds to the weighted calculation of [Precision, Recall, mAP@0.5, mAP@0.5:0.95], and then finding the maximum value to determine the best weights. For each pre-trained model, YOLOv5s and YOLOv5m, after completing 200 iterations of model training, we had four weights: s_best.pt, s_last.pt, m_best.pt, and m_last.pt. These weights correspond to the best.pt and last.pt obtained from training the two pre-trained models.

To perform the final performance evaluation, we used the dataset's test set. We tested the best.pt obtained from each training session on machines A and C to determine the real-time detection capabilities of the same model in different configurations. Secondly, we tested the four obtained weights on Machine C to assess the model's target detection accuracy under the same configuration. The relevant test data obtained are shown in Table 6 and Table 7. According to the results, both YOLOv5s and YOLOv5m can perform real-time object detection on devices with GPUs. The best.pt, and last.pt obtained using YOLOv5m as the pre-trained model have similar accuracy, slightly higher than those obtained using YOLOv5. However, YOLOv5m has a larger number of parameters, and its model size is three times that of YOLOv5s.

Table 6: Test Performance

Model	Size(MB)	Machine	Inference Time(ms)
s_best.pt	**14.49**	A	279.00
		C	**3.3**
m_best.pt	42.33	A	640.0
		C	**7.4**

Table 7: Test Performance

Model	Average Precision (All)	Average Precision (Visona Only)
s_best.pt	**0.81**	**0.90**
s_last.pt	**0.82**	**0.90**
m_best.pt	0.84	0.92
m_last.pt	0.84	0.92

4.5 Conclusion

Based on the project's specific requirements and resources, the best.pt uses the YOLOv5m model obtained from training was chosen for the decision-making layer design. This model achieves a detection time of only 3.4ms on machines with GPU configuration. It successfully performs real-time object detection for 23 object classes with an accuracy of 84%. The average accuracy for detecting tactile paving, zebra crossing, and steps reaches 92%. The model meets the accuracy and real-time performance requirements, providing a safe and reliable assistive solution for visually impaired individuals.

5. Hardware Design and Prototyping

By further applying the trained YOLOv5 road condition detection model, the assistive device design for visually impaired can be optimized to maximize the accuracy of information informing and ensure outdoor safety while minimizing changes to users' habits to reduce cognitive load. Considering the stakeholders and the idea of sustainable design, the use of technology should be balanced to avoid

ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

unnecessarily high costs. From this perspective, the Arduino platform was chosen for development instead of more powerful platforms like the RaspberryPi. A successful design framework was constructed, and a device prototype was implemented using Arduino development technology. The main body of the device can be flexibly installed on any equipment like a white cane or simply held by a hand, and it incorporates visual information into auditory and tactile feedback. The framework design consists of three parts: the decision-making layer, communication layer, and execution layer, as shown in Figure 3.

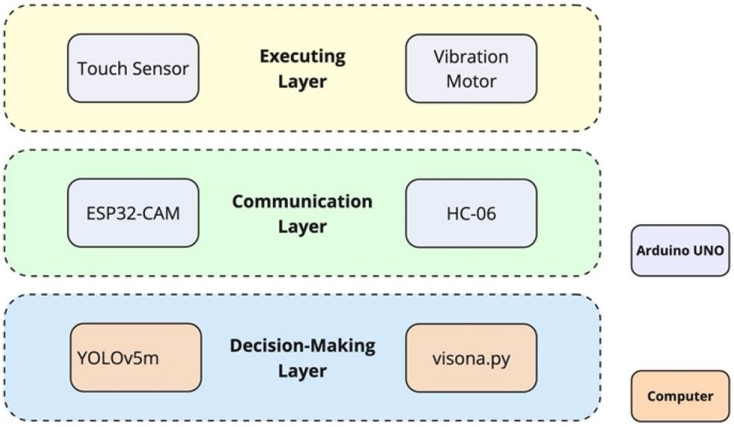

Figure 3: Design Framework

The project's functional design includes functions as follows:

(1) Environmental information feedback: The environmental information feedback enables the detection of objects at longer distances, and the corresponding interactive feedback provides users with more information about the detected objects, including their category and position.

(2) Tactile paving search mode: When users deviate from the tactile paving, they can activate the tactile paving search mode by pressing a button. The device then performs target detection only for the tactile paving category and triggers vibration feedback in the interaction module.

(3) Privacy intrusion warning: The system is designed to detect the "people" category in the object detection results. When someone is detected too close to the user, the system utilizes face recognition to identify the person and generates corresponding interactive voice prompts using gTTS. The interaction with the user is achieved through controlling the vibration motor and voice prompts. This feature gives users a certain level of social initiative and enhances their awareness of the environment, including people, enabling their ability to take self-protective measures beforehand.

To implement the designed functionalities, the project makes full use of the hardware interfaces on Arduino UNO R3, and the main sensors and related interactive hardware involved are shown in Figure 4. below:

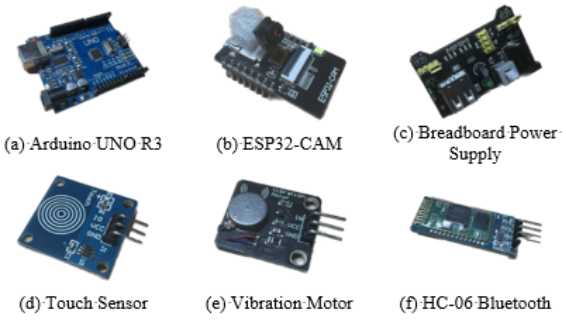

Figure 4: Hardware Interfaces

With the successful design and implementation of the framework, the project provides users with more significant response space and ensures safer travel while minimizing the device's "presence." The final prototyping design is shown in Figure 5. Taking sustainability into consideration, the cost has been kept to a minimum. By integrating the project document on the Calob platform and Github, practitioners can continuously expand the existing design framework based on local conditions, allowing it to benefit from the open design and contribute to its future development.

ISSN 2616-5775 Vol. 6, Issue 10: 106-112, DOI: 10.25236/AJCIS.2023.061016

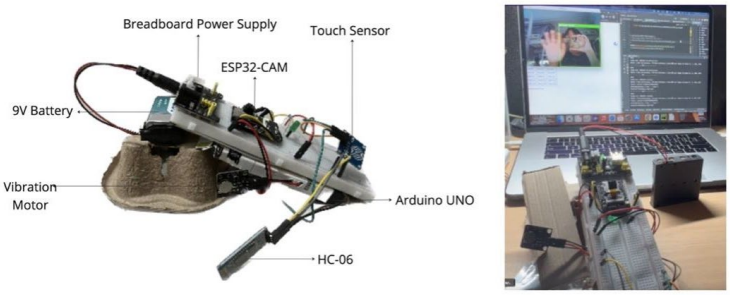

Figure 5: Prototyping

6. Conclusion and Prospective

This project's proposed optimization design framework for assistive devices for visually impaired individuals utilizes the YOLOv5 algorithm for real-time object detection of 23 target classes. Considering both accuracy and inference time, the YOLOv5m trained model was selected as the core of the decision-making layer. With the development using Arduino, a prototype was successfully implemented. By further exploring, it is found that devices equipped with Bluetooth or WiFi modules allow object detection to perform on devices with GPUs, and the corresponding detection results can be transmitted to the execution layer through communication technology. In theory, this design framework can fully meet the practical standards regarding safety and accuracy.

However, since the project did not optimize the model, the obtained weight is not considered outstanding compared to other design outcomes mentioned in the literature review, leaving ample room for optimization. Additionally, there are several areas for improvement in the application of the algorithm. Firstly, the current prototype is primarily intended for demonstration and temporarily depends on a computer. However, it is possible to iterate towards a completely wireless mobile device by utilizing the ESP32-CAM for communication with Python via a web server. Secondly, systematic designs for nighttime travel and transportation scenarios have yet to be developed. Thirdly, later development will consider providing visually impaired individuals with more proactive obstacle avoidance functionality rather than solely relying on passive avoidance. These are potential research directions for the project in the future.

Looking ahead, drawing inspiration from outstanding global innovative design cases, it is envisioned to build an ecosystem centred around Visona and incorporate it into social innovation design. With its robust scalability and low technical barriers, efforts can be made to balance the various stakeholders in the ecosystem and make Visona democratic and accessible to the public.

References

[1] Wu J. Introduction to convolutional neural networks [J]. National Key Lab for Novel Software Technology. Nanjing University. China, 2017, 5(23): 495.
[2] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, United States: IEEE Computer Society Conference Publishing Services, 2016: 779-788.
[3] Chen D, Gao Y, Song A, Liu J, Zeng H. Touchscreen interactive finger-mounted Braille reproduction system[J]. Chinese Journal of Scientific Instrument, 2022, 43(5): 199-208.
[4] Sáez Y, Montes H, Garcia A, et al. Indoor navigation technologies based on RFID systems to assist visually impaired people: A review and a proposal [J]. IEEE Latin America Transactions, 2021, 19(8): 1286-1298.
[5] Du Y, Yuan X, Ma X. Design of an Intelligent Cane Based on Multi-sensor Fusion Technology [J]. Practical Electronics, 2021(07): 80-81+19.
[6] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M. Kitani, et al. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People[C]//Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19). New York, New York, United States: Association for Computing Machinery, 2019:68–82.
[7] Buchanan R. Design research and the new learning [J]. Design Issues, 2001, 17(4): 3-23.
[8] Brown T, Wyatt J. Design thinking for social innovation [J]. Development Outreach, 2010, 12(1): 29-43.