Traffic & Transportation

Intelligent Obstacle Avoidance Method for Blind Travel Based on an Improved YOLO Algorithm and Binocular Vision

Yan ZHUANG[1]

Original Scientific Paper
Submitted: 17 June 2025
Accepted: 18 Sep 2025
Published: 27 May 2026

[1] zhuangyan123452024@163.com, School of Information Engineering, Anhui Business and Technology College, Hefei, China; School of Computer Science and Technology, Anhui University, Hefei, China

This work is licensed under a Creative Commons Attribution 4.0 International Licence.

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT
With the improvement of deep learning technology, traditional obstacle avoidance approaches for blind people can no longer meet practical needs. In response to the problems of poor adaptability to multiple scenarios and low obstacle avoidance rates in current obstacle avoidance methods for blind people, this study proposes an intelligent obstacle avoidance method for blind people based on an improved You Only Look Once (YOLO) algorithm and binocular vision. This method first improves the YOLOv8 algorithm by introducing an angle loss function, SPD convolution and a more efficient local convolution structure, and proposes an I-YOLOv8-BBOD model suitable for blind people's travel. Then, based on this model, a blind travel intelligent obstacle avoidance platform is built. The confusion matrix analysis showed that the research model performed the best and had significant advantages in reducing false positives and false negatives. Its precision, recall and F1 score were all above 0.90, indicating the best overall performance. The ablation experiment showed that after the improvement of each module, the mAP and the average accuracy at the 50% IoU threshold increased by 13.19% and 16.00%, significantly improving the detection accuracy. In the example application, the successful obstacle avoidance rate of the proposed obstacle avoidance platform exceeded 90%, and the false alarm rate was below 1.3%. The traditional platform could reach up to 5.4%, which was better than the previous platform. This indicates that the research method can accurately detect obstacles and road conditions, ensuring the safety of visually impaired people's travel.

KEYWORDS
YOLOv8; blind people; obstacle; binocular vision; intelligent obstacle avoidance; lightweight.

1. INTRODUCTION

With the promotion of artificial intelligence and computer vision (CV) technology, their achievements have been deeply integrated into industrial inspection, autonomous driving, smart healthcare and other scenarios, reshaping production and lifestyle with efficiency and accuracy beyond human perception [1-2]. However, the existing technological system is mostly built based on the visual cognition of a healthy population, ignoring the special needs of visually impaired groups for environmental perception. Traditional auxiliary equipment, such as ultrasonic waveguide blind canes, has limitations in detection range and cannot identify obstacle types, making it difficult to meet the safety navigation needs in complex urban environments [3]. Currently, the integration of CV and embedded technology has brought breakthrough opportunities to the field of visually impaired assistance. With the breakthrough in computing power of system on chip (SoC) for smartphones and the introduction of lightweight neural networks, it has become possible to run deep learning (DL) models in real-time on mobile devices. Binocular vision technology can obtain centimetre-level depth information through disparity calculation, and when combined with object detection algorithms, it can achieve obstacle

spatial localisation and semantic understanding [4]. However, there are still three issues with existing research. Firstly, the general object detection model lacks the ability to distinguish small obstacles (such as manhole covers and road stakes) and highly similar objects (such as shared bicycles and fire hydrants) in visually impaired scenes. Secondly, the robustness of binocular matching algorithms under complex lighting conditions (strong direct light, low light shadows) is weak. Thirdly, there is a lack of end-to-end system architecture optimised for interaction with visually impaired users, which makes it difficult to convert detection results into actionable obstacle avoidance instructions [5-6]. Therefore, many experts have researched issues in this field. To address the instability issues in current single-sensory braille guidance methods, Rui et al. developed a multimodal braille guidance system combining YOLO and ORB-SLAM, delivering tactile and auditory feedback. The results demonstrated real-time accuracy in detecting obstacle locations and categories within environments. By integrating YOLO and ORB-SLAM, this system provides a practical assistive device for the visually impaired community, enabling safe navigation [7]. For narrow and crowded pedestrian areas where conventional obstacle avoidance methods struggle, Mu et al. proposed a dynamic obstacle avoidance system based on a fast instance segmentation network (FISN). Experimental results showed that this system effectively assists visually impaired users in safely navigating challenging environments [8].

To reduce the obstacle interference that autonomous underwater vehicles (AUVs) may encounter during mission execution, Cao et al. conducted research on sonar-based obstacle detection (OD) and avoidance methods for AUVs. The proposed algorithm has been found to improve the accuracy of OD and the processing speed of sonar images. Meanwhile, the algorithm ensured the safe navigation of AUVs in complex obstacle environments [9]. To improve the OD rate on the path of unmanned aerial vehicles (UAVs), Rahman et al. used UAV aerial photography to achieve real-time OD of You Only Look Once version 8 (YOLOv8) in wireless sensor networks. The YOLOv8 outperformed other baseline algorithms in various scenarios, and the F1 score reached 96% within 200 cycles [10]. To improve the efficiency of precision agriculture management, Wang et al. proposed a UAV perspective OD method based on a non-local deformable detection transformer. The mean average precision (mAP) value has increased from 71.3% to 78.0% [11]. To enable unmanned agricultural vehicles to detect various types of obstacles in farmland in real-time and accurately, Xue et al. proposed an improved YOLOv5s built on the K-means algorithm and CIoU function to lift detection accuracy and accelerate timely detection speed. The mAP has increased by 5.80% in comparison with the original YOLOv5s, indicating that utilising the CIoU significantly reduced missed and false detections in the previous YOLOv5s [12]. To improve the perception ability of autonomous robots towards dynamic obstacles, Xu et al. proposed a lightweight 3D dynamic OD and tracking method based on RGB-D cameras. Flight experiments have shown that the tracking results can effectively change the trajectory of the robot to navigate in dynamic environments [13]. To effectively solve the problem of visual impairment, Suman et al. combined a single detection mechanism with recursive neural networks to construct an intelligent obstacle recognition model for visually impaired users. The best accuracy calculated by this model for indoor and outdoor scenarios was 95.06% and 87.68%, respectively. The accuracy of the model for close and long-distance obstacles was 96.4% and 86.8%, which was superior to other models. Considering all common obstacles, the execution time of the model was 4.82 seconds, and the highest average accuracy was 95.54% [14].

In summary, CV technology has achieved significant results in OD. In multiple fields, improved algorithms have improved detection accuracy, speed and precision. For example, algorithms based on sonar, YOLO, etc., have shown excellent performance in AUVs, drones, unmanned agricultural vehicles and other scenarios, with some models having high accuracy. However, these studies still have shortcomings, such as the long execution time of models for visually impaired users and the lack of comprehensive solutions to complex dynamic environments, multi-scene adaptability and other issues. Based on this, this study innovatively proposes a Blind Travel Intelligent Obstacle Avoidance Platform (BTIOAP) for visually impaired people from the perspective of travel, based on an improved YOLO algorithm and binocular vision. This study first improves the loss function of the YOLOv8 network by introducing the SIOU function. At the same time, the weight calculation method is adjusted using the SlideLoss function to solve the problem of low recall and precision in YOLOv8 object detection. Then, this study introduces SPD-conv and a more efficient partial conv (PConv) structure to improve the backbone layer, to address the high computational resource requirements of YOLOv8. Finally, based on the improved YOLO algorithm, this study constructs a BTIOAP. This study aims to break through the limitations of existing technology in adapting to visually impaired scenarios, promote the value extension of CV technology from “assisting able-bodied people” to “empowering disabled people”, and provide technical support for building a universal and accessible society.

2. METHODS AND MATERIALS

2.1 Design of loss function for I-YOLOv8-BBOD algorithm

At present, computer object detection technology is widely used in various fields. However, with the advancement of science and technology, there are still shortcomings in the application of current computer detection technology in guide devices, such as a high false detection rate in complex travel environments and weak real-time semantic understanding ability [15]. In response to the problems existing in current object detection algorithms, this study combines the Improved YOLOv8 with the binocular vision object detection algorithm from aspects such as PConv and loss function, proposing an I-YOLOv8-BBOD model suitable for blind people's travel.

The traditional YOLOv8 algorithm adopts a three-stage structure. The backbone layer extracts features through convolution and transpose convolution operations, and introduces residual connections, bottleneck structure modules, and spatial pyramid pooling (SPPF) modules to optimise performance. Among them, the residual module includes a convolution module, a cross-stage partial bottleneck with 2convolutions (C2f). The neck layer combines multi-scale features through upsampling and path aggregation network (PAN) components to enhance feature expression. The head layer adopts a separated design. The classification head and detection head independently optimise the classification loss (VFL loss) and localisation loss (DFL Loss+IoU loss), and finally fuse the two types of losses to calculate the total loss to guide model training. The function is shown in *Equation 1*.

$$CIoU = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{1}$$

In *Equation 1*, b and b^{gt} are the centre points of the predicted box and real box. c denotes the diagonal distance between the closed areas of two rectangular boxes. ρ means the calculated Euclidean distance, α indicates the weight coefficient. v is the consistency of measuring the ratio between the width and height of two rectangular boxes. However, the use of CIoU loss is computationally complex and introduces multiple factors, resulting in increased computational overhead, especially in large-scale datasets and complex models, which may significantly increase training time. In addition, when the discrepancy in aspect ratio between the predicted and true boxes is too big, the values of the related terms will abnormally increase [16-17]. For this issue, this study improves the object detection function of the YOLOv8 algorithm. Firstly, this study introduces the angle loss function (shape-IoU, SIOU). The angle loss Ω of this function is described in *Equation 2*.

$$\Omega = 1 - 2 \times \sin^2\left(\arcsin\left(\frac{h}{\chi}\right) - \frac{\pi}{4}\right) \tag{2}$$

In *Equation 2* h is the height difference between the actual and predicted box centre points. χ is the distance between two points. The distance loss ∂ of the SIOU function also considers the angle loss, as shown in *Equation 3*.

$$\partial = \sum_{t=x,y} 1 - e^{-\gamma \rho_t} \tag{3}$$

In *Equation 3*, ρ_t and w are the square value and width of the minimum bounding rectangle width between the actual and predicted box, $\gamma = 2 - \Omega$. h_1 is the height of the rectangle. $\left(o_{ax}, o_{ay}\right)$ and $\left(o_{px}, o_{py}\right)$ are the coordinates of the centre point of the actual and predicted box. The description of the shape loss ℓ of the SIOU function is shown in *Equation 4*.

$$\ell = \left(1 - e^{-w_1}\right)^{\theta} + \left(1 - e^{-h_2}\right)^{\theta} \tag{4}$$

In *Equation 4*, w_1 and h_2 are the width and height of the prediction boxes. θ is the correlation of shape loss, within the range of $[2,6]$. This range is determined based on previous experience [18]. Finally, the loss description of the SIOU function is shown in *Equation 5*.

$$L = 1 - IOU + \frac{\partial + \ell}{2} \tag{5}$$

This study can replace the original CIoU loss of YOLOv8 with the SIoU loss to improve the convergence speed and precision. To address the issue of detecting more details or smaller target categories in practical applications, to improve the performance, this study further focuses on the SlideLoss function and makes detailed and meticulous adjustments to its weight calculation method. This adjustment measure aims to guide the algorithm network to be more inclined to learn challenging samples. The weight allocation scheme of the SlideLoss function is shown in *Figure 1*.

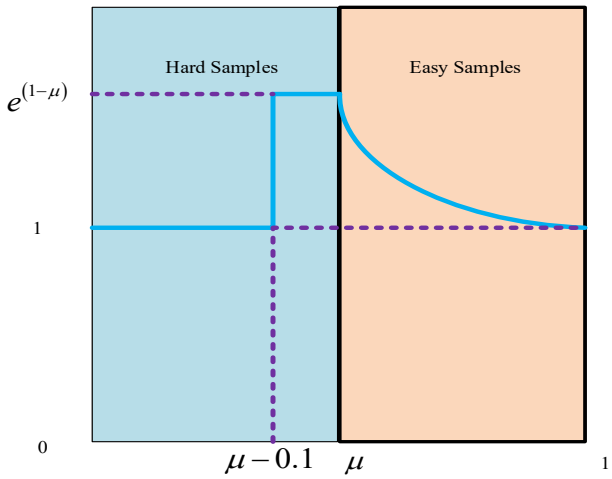

Figure 1 – Weight allocation scheme for the SlideLoss function

The SlideLoss function takes the average IoU value of all bounding boxes as the threshold μ, which depends on the characteristics of the entire dataset and is dynamically adjusted according to changes in the data. For samples in object detection tasks, this study classifies them based on their IoU values with real boxes for ease of classification. Specifically, based on a preset threshold, samples with IoU values below the threshold are considered difficult samples and given greater weights to enhance their contribution in loss calculation. Samples above the threshold are considered simple samples with lower weights. For decision boundary region samples that are prone to misclassification, the weights are further increased to enhance the classifier's attention to difficult samples, thereby optimising the overall classification performance. However, when the IoU value of the sample is below a specific threshold $[\mu, \mu - 0.1]$, it is judged as an extremely difficult or false positive (FP) sample and its weight is not adjusted. The SlideLoss strategy effectively improves the IoU value and detection precision of the model by prioritising the learning of difficult samples, enhancing the target recognition ability in complex scenes.

2.2 PConv design of the I-YOLOv8-BBOD algorithm

After optimising the loss function of the YOLOv8, the next step is to further enhance the computational efficiency of the YOLOv8 to better meet the stringent demands for real-time and high efficiency in practical applications. Therefore, after repeated research and demonstration, this study introduces SPD-Conv and the PConv structure with superior performance and higher efficiency to comprehensively and meticulously improve the backbone layer. The backbone network of traditional YOLOv8 downsamples and enlarges the receptive field through a convolution module with a stride of 2, and generates multi-scale feature maps. However, this process will lose detailed information, especially affecting the performance of small target detection. Multiple operations in the convolutional layer can easily dilute the features of small targets, and the convolutional structure is sensitive to background noise, which can lead to FPs or false negatives (FNs) in complex backgrounds. To address this issue, this study introduces a novel convolution module called SPD-Conv in the primary downsampling phase to replace the previous convolution layer with a stride of 2. Its structure is shown in *Figure 2*.

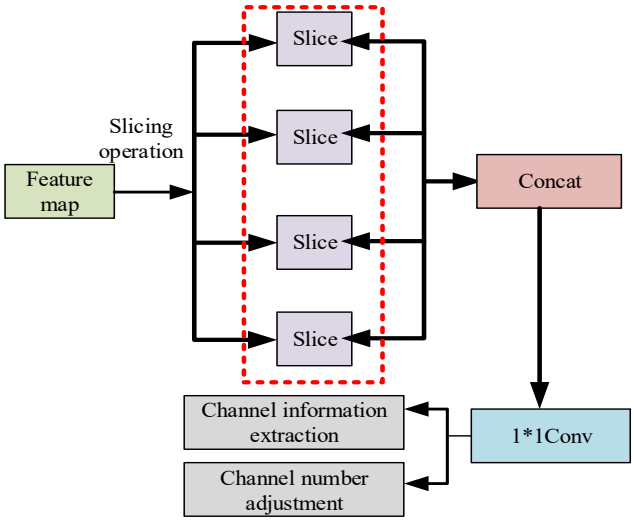

Figure 2 – SPD-Conv framework

In *Figure 2*, the SPD-Conv structure starts from the input feature map and divides it into multiple slices through slicing operations. These slices are then concatenated together to form a new feature map. Next, the concatenated feature maps are processed using 1×1 convolution to adjust the number of channels or fuse channel information. Meanwhile, the structure also includes channel information extraction and channel number adjustment modules for further processing of channel information. This design enhances the feature extraction capability and flexibility of the model. The C2f module is the core component of the backbone layer, which relies on multiple layers of DarknetBottleneck modules to deeply explore the features of the feature map. Considering the redundant computing problem caused by this method, this study introduces PConv to improve the C2f module, and the improved module is called C2f-PC, with its structure shown in *Figure 3*.

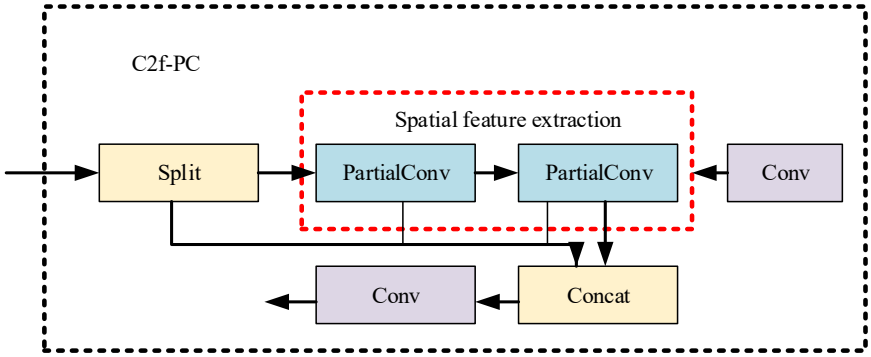

Figure 3 – C2f-PC structure

In *Figure 3*, the input feature map is first split by a split operation, and then the PConv layer extracts spatial local features through PConv. Subsequently, the output of multi-level PConv layers is concatenated with the results of conventional convolutional layers, and multi-level features are fused to enhance the performance and precision of YOLOv8. PConv reduces computational complexity by restricting the number of computation channels, minimising floating point operations (FLOPs) and memory access, especially for devices with limited I/O resources [19-20]. Although only partial channels are processed, the selected channel information can still be circulated and utilised through subsequent point-by-point convolutional layers to ensure feature integrity.

2.3 Construction of BTIOAP

After completing the PConv and loss function design of the YOLOv8 algorithm, this study finally obtains an I-YOLOv8-BBOD model suitable for blind people's travel, as shown in *Figure 4*.

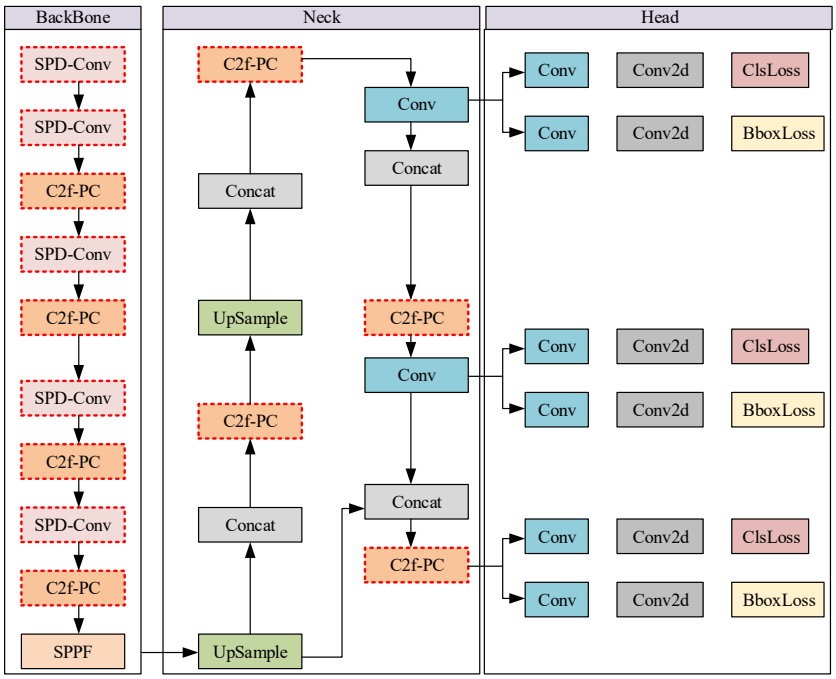

Figure 4 – I-YOLOv8-BBOD model architecture diagram

In the proposed model, SPD-Conv can achieve downsampling without sacrificing information integrity, thereby enhancing the accuracy of small object detection. In contrast, PConv utilises redundant information in feature maps to reduce computational burden and memory access requirements. Specifically, PConv only applies regular convolution to a portion of the input channel to extract spatial features, while keeping the remaining channels unchanged. Finally, based on the binocular vision detection model, this study develops a BTIOAP with real-time detection of common road objects, voice interaction and map navigation functions. The architecture of the platform is shown in *Figure 5*.

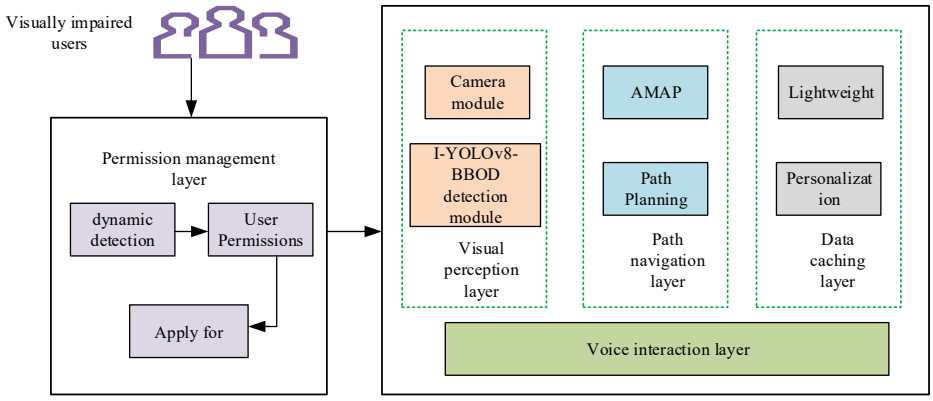

Figure 5 – Overall architecture design of BTIOAP

In *Figure 5*, the platform is a modular architecture with user interaction and scene perception as its core, covering six functional units: permission management, visual perception, intelligent decision-making, spatial navigation, voice interaction and data persistence. The permission management module serves as the underlying support layer, providing dynamic detection and one-click application capabilities for sensitive permissions such as camera calling, location acquisition and voice control, ensuring compliance. The visual perception layer is jointly driven by the camera module and the binocular vision detection module. The former completes camera parameter initialisation and real-time preview flow control. The latter converts the input image into structured risk text using the I-YOLOv8-BBOD model. The textual process of its target recognition results is shown in *Figure 6*.

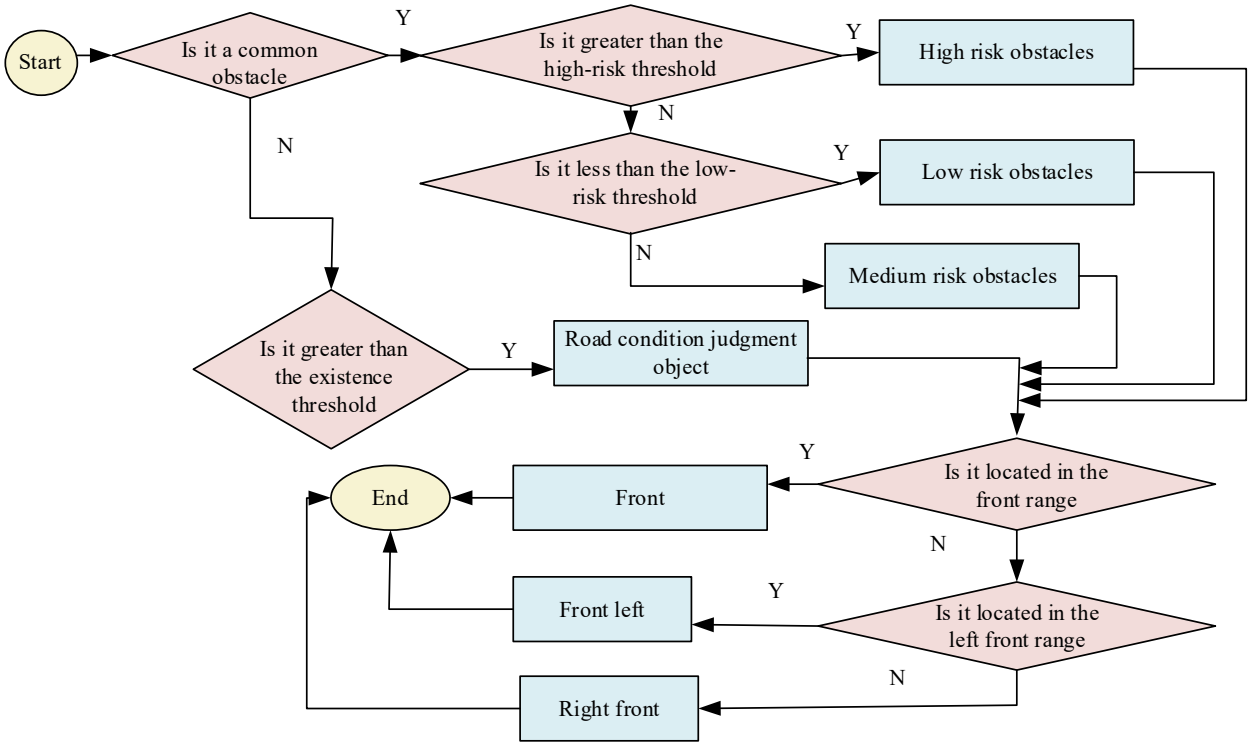

Figure 6 – Text-based flowchart of target recognition results

This process optimises the ability to parse environmental information for blind people's travel scenarios. The first step is to standardise the image data collected by the camera, and adjust the image to a size of 640×640 through a scaling matrix. The next step is to pass it through the JNI interface to the C++ environment for model inference, and output a Float array containing target coordinates, categories and confidence levels. After eliminating redundant detection boxes through NMS, the inference results are mapped to coordinate space to generate the final recognition result, which supports real-time image annotation or textual output. In response to the needs of blind users, this study establishes a risk classification mechanism. This mechanism is based on setting low, medium and high risk thresholds for target size, and setting independent existence thresholds for different obstacles. It determines the existence and risk level of an object by comparing its size with a threshold, and combines coordinate mapping orientation information to ultimately generate a text prompt in the format of "description+orientation+category".

The spatial navigation module is based on the Amap SDK and provides functions such as geographic positioning, fuzzy search of points of interest and pedestrian path planning. It supports quick access to navigation services through voice commands or text input [21]. The voice interaction layer achieves full scene coverage through the iFlytek engine, customises voice command sets for various functional interfaces, and supports braille voice broadcasting and multi-round dialogue-based operation guidance. The cache module serves as a local data hub, managing user personalised settings with a lightweight key-value storage solution to ensure consistency in cross-device experience.

3. RESULTS

3.1 Dataset establishment and performance testing of the I-YOLOv8-BBOD model

To validate the effectiveness of the research method, a self-built WOTR dataset is used for performance testing of the I-YOLOv8-BVOD model. This study first focuses on the travel scenarios of visually impaired individuals and determines the recognition target category of the dataset through scenario analysis. In response to the needs of visually impaired people to avoid obstacles and cross roads during travel, the targets are divided into two categories: common obstacles and road condition judgment objects. The outdoor survey found that the tactile paving is often occupied by vehicles, roadblocks, etc., which seriously affects the travel of the visually impaired. Therefore, it is necessary to focus on screening 15 types of common obstacles around the tactile paving and sidewalk, and subdivide them into static (such as trees, garbage cans), dynamic (such as

pedestrians, pet dogs) and unknown states (such as bicycles parked and running). This study emphasises the need to pay more attention to static and unknown obstacles, as they pose a significant threat to the mobility of visually impaired individuals. At the same time, this study selects five types of road condition judgments, including traffic lights (red lights, green lights), crosswalks, tactile paving and signs. These objects can help visually impaired individuals confirm their location and road conditions, assisting them in safely crossing the road. Ultimately, the constructed WOTR dataset includes 15 common obstacles (7 static, 2 dynamic, 6 unknown) and 5 road condition judgment objects to accurately reflect real-world application scenarios. Finally, through filtering and data annotation, the WOTR dataset produced contains a total of 13,928 images, with 9,056, 2,534 and 2,338 images in the training, testing and validation sets. The entire dataset contains nearly 190,000 bounding boxes. The operating environment and parameter settings are listed in *Table 1*.

Table 1 – Experimental operating environment and parameter settings

/	**Category**	**Set items**
JRE	Operating system	Win 7 operating system
	Developer components	JDK 7, Eclipse
	CPU	Intel (R) Core (TM) i9-10900X
	GPU	Nvidia RTX 3090 GPU
	Platform software environment	Python3.8, PyTorch1.10.0
	Programming language	Java
Parameter setting	Batch size	8
	Learning rate initial value	0.001
	Learning rate initial value	Adam

This study compares single-stage detection models based on focal loss (RetinaNet), end-to-end object detection models based on transformer architecture (DETR), and traditional YOLOv8 as research algorithms. Precision, recall rate and F1 score are selected as evaluation indicators for the experiment. Firstly, a comparison of the iterative performance of each model is conducted on the training and testing sets of the self-built dataset WOTR, as displayed in *Figure 7*.

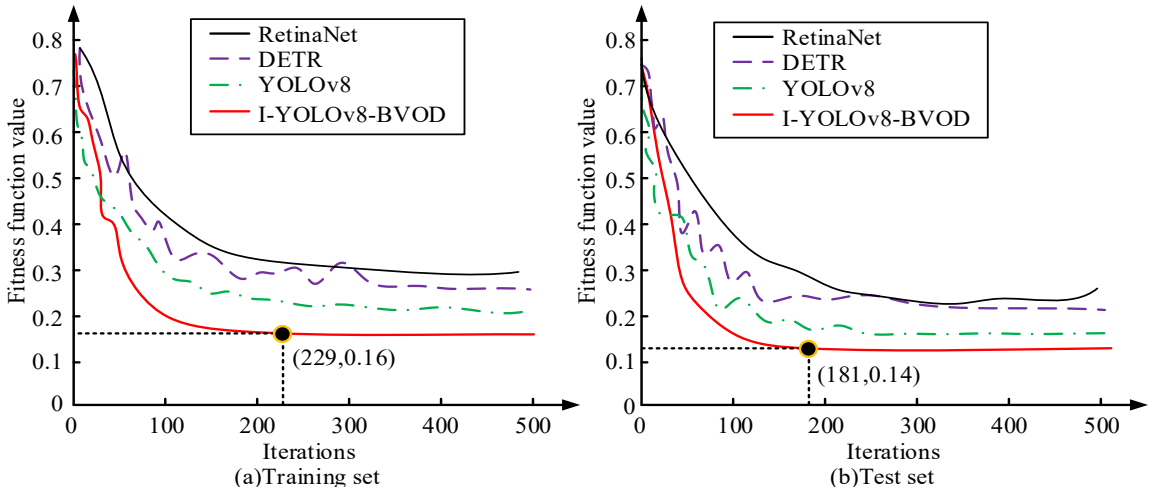

Figure 7 – Iterative performance of four models in two sets

Figures 7 (a) and *(b)* show the fitness iteration curves of different models in two datasets. I-YOLOv8-BBOD has the best iteration performance in both the training and testing sets. In the training set, I-YOLOv8-BBOD only needs 229 iterations to reach a stable state, with a fitness value of 0.16. In the test set, the model can reach a stable state after 181 iterations, and the fitness value in the stable state is 0.14. This indicates that I-YOLOv8-BBOD is effective and can be used for obstacle and road condition detection. *Table 2* displays the confusion matrix data of various models in the test set of the nuScenes dataset.

Table 2 – Analysis of confusion matrix for different detection models

Different detection models		Actual positive example	Actual positive example	Precision	Recall	F1 scores
RetinaNet	Predict positive examples	TP=800	FP=200	0.80	0.84	0.82
	Predict the counterexample	FN=150	TN=1850			
DETR	Predict positive examples	TP=750	FP=250	0.75	079	0.77
	Predict the counterexample	FN=200	TN=18D0			
Traditional YOLOv8	Predict positive examples	TP=850	FP=150	0.85	0.89	0.87
	Predict the counterexample	FN=100	TN=1900			
I-YOLOv8-BVOD	Predict positive examples	TP=900	FP=100	0.90	0.95	0.92
	Predict the counterexample	FN=50	TN=1950			

In *Table 2*, RetinaNet performs moderately, with high FP and FN, indicating that the model has certain issues with FPs and FNs. DETR performs poorly, while FP and FN are both high, indicating limited performance of the model in object detection tasks. Traditional YOLOv8 performs well with low FP and FN, indicating that the model has high accuracy in object detection tasks. I-YOLOv8-BBOD performs the best, while FP and FN are the lowest, indicating that the improved model has significant advantages in reducing FP and FN. In addition, the I-YOLOv8-BBOD model has significant advantages, with a precision of 0.90, which is higher than other models, indicating a lower FP rate. The recall rate is 0.95, higher than other models, indicating a lower FN rate. The F1 score is 0.92, indicating the best overall performance and achieving a better balance between precision and recall. This is due to the improved methods introduced, which enable it to maintain high detection accuracy and robustness in complex scenes. This ablation experiment uses the YOLOv8 network as a benchmark and tests it on a test set of a self-built dataset, as shown in *Figure 8*.

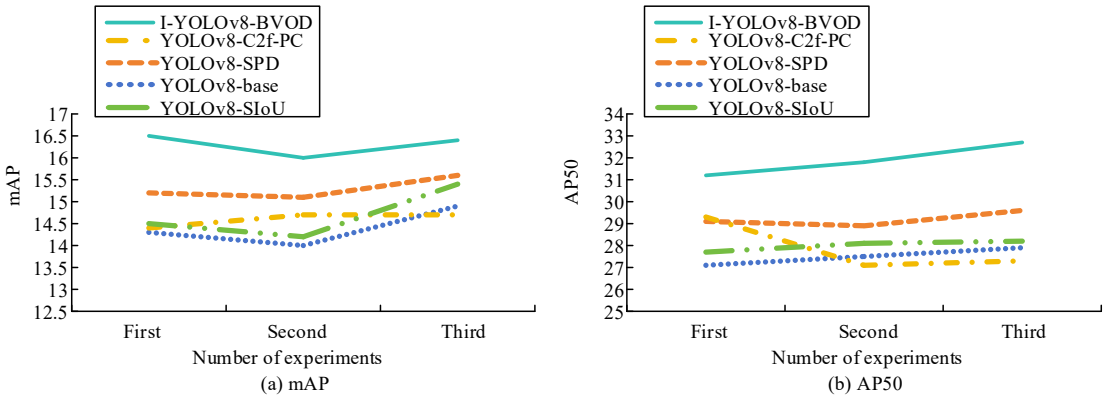

Figure 8 – Experimental results of the I-YOLOv8-BBOD model ablation

In *Figure 8*, after introducing the SPD-Conv structure, mAP and AP50 improve by 6.25% and 6.18% compared to the base model, effectively solving the problem of decreased detection precision caused by information loss in the benchmark model. After using C2f-PC, mAP and AP50 improve by 1.39% and 1.45% compared to the basic models. This module can enhance the C2f's ability to learn and express features, and add a small object detection layer to improve the detection accuracy of small objects. After improving the loss function, the mAP and AP50 increase by 2.08% and 1.82% compared to the basic model, which can speed up the convergence velocity of the model. After improving each module, the mAP and AP50 of I-YOLOv8-BBOD increase by 13.19% and 16.00% compared to the basic model. I-YOLOv8-BBOD can significantly

improve the detection accuracy of obstacles and road conditions. To further validate the applicability performance of I-YOLOv8-BBOD, this study conducts performance tests on the ImageNet and COCO datasets, and the results are shown in *Figure 9*.

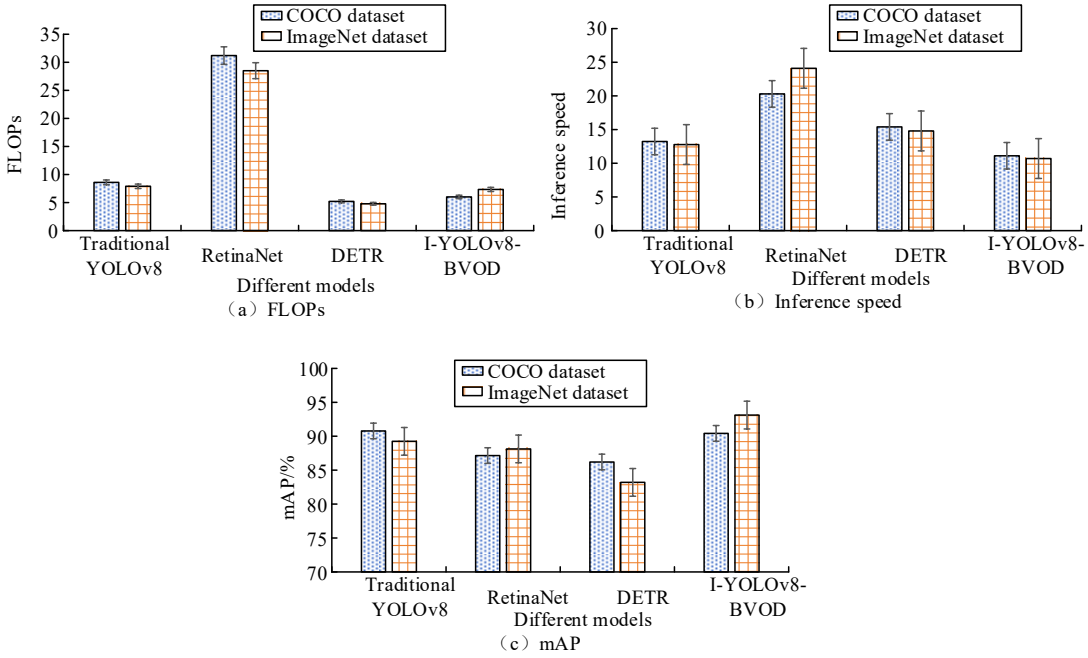

Figure 9 – Model testing results under different datasets

Figures 9 (a) – (c) show the test results of FLOPs, inference speed and mAP for different algorithms. In *Figure 9 (a)*, the FLOPs of the research algorithm on the COCO and ImageNet datasets are 6.0 and 7.3, which are better than the comparison algorithm, with improvements of 30.2% and 7.09% compared to YOLOv8. In *Figure 9 (b)*, the inference speed of the research algorithm on two datasets is 11.1s and 10.7s, which is 16.0% and 16.2% faster than the YOLOv8 algorithm. In *Figure 9 (c)*, the mAP of the proposed algorithm on the dataset is 90.7% and 93.1%, which is 14.3% and 13.2% higher than traditional YOLOv8. This indicates that the research model can significantly improve detection speed while enhancing detection accuracy, achieving a balance between the two and making obstacle and road condition detection more effective and reliable.

To evaluate the performance of extreme scenario models, the study selected four scenarios (direct strong light exposure, low-light conditions during rainy nights, crowded rush hour traffic and cluttered construction zones). Each scenario was supplemented with 500 images for the WOTR dataset. Comparative tests were conducted using the I-YOLOv8-BVOD model against traditional YOLOv8 models, with evaluation metrics including mean average precision (mAP), recall rate and inference speed. The test results for extreme scenario models are presented in *Table 3*.

Table 3 – Performance test results of the extreme scenario model

Scenario type	Model	mAP (%)	Recall (%)	Inference speed (s)
Strong light direct	I-YOLOv8-BVOD	82.3	85.1	12.5
	YOLOv8	68.7	70.2	13.8
Low light levels at night in the rain	I-YOLOv8-BVOD	79.5	81.3	13.1
	YOLOv8	65.2	67.4	14.5
Highly populated	I-YOLOv8-BVOD	84.7	88.2	11.8
	YOLOv8	72.1	75.3	12.9
Disorganised construction area	I-YOLOv8-BVOD	80.1	83.5	12.7
	YOLOv8	66.9	69.8	14.2

As shown in *Table 3*, the I-YOLOv8-BVOD model outperforms traditional YOLOv8 in extreme scenarios, with mAP and recall rates outperforming by an average of 12.8% and 13.5%, respectively, while achieving an average inference speed improvement of 1.2 seconds. In strong-light and rainy/night environments, SPD-Conv's feature retention capability effectively mitigates illumination-induced feature loss. For crowded pedestrian zones and chaotic construction sites, the SIOU loss function optimises target bounding box localisation accuracy while reducing missed detections of overlapping targets. However, model performance still shows an 8%–10% decline compared to stable scenarios under extreme conditions. Future improvements should focus on enhancing robustness through data augmentation (e.g. illumination perturbations, rain/mist simulations) and attention mechanism optimisation.

3.2 Application analysis of BTIOAP instance

After testing the performance of the I-YOLOv8-BBOD model, this study further conducts instance application testing on the proposed BTIOAP. This study deploys the BTIOAP for daily travel scenarios of visually impaired users in an old community in a certain city (with an average road width of 3.2 metres and high obstacle density). In a continuous week of testing, the successful obstacle avoidance rate and false alarm rate (FAR) of the BTIOAP compared to traditional platforms on community sidewalks are recorded, as shown in *Figure 10*.

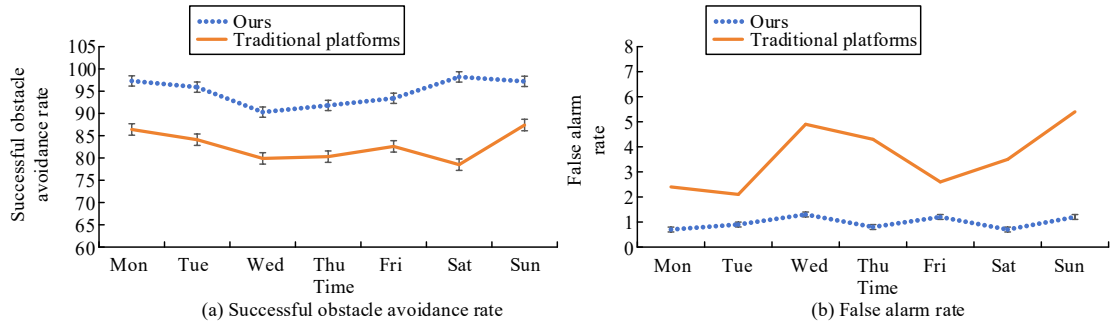

Figure 10 – BTIOAP test results

Figures 10 (a) and *(b)* show a comparison between the platform's successful obstacle avoidance rate and the platform's FAR. In *Figure 10 (a)*, the successful obstacle avoidance rates of the proposed BTIOAP are all above 90%, while the traditional platform only reaches a maximum of 87.4% during the testing period. In *Figure 10 (b)*, the FAR of BTIOAP is consistently below 1.3%, while the traditional platform reaches a maximum of 5.4% during testing. This result indicates that BTIOAP can accurately detect obstacles and road condition judgment objects, helping to provide safe and efficient travel solutions for visually impaired groups.

Finally, 15 visually impaired users (8 completely blind and 7 with low vision) were invited to participate in the two-week field test of the platform. Feedback was collected through questionnaires (Likert 5-point scale) and semi-structured interviews, and the evaluation dimensions included functional practicability, operational convenience, information accuracy and sense of security. The evaluation results are shown in *Table 4*.

Table 4 – Feedback results of visually impaired users

Assessment dimensions	Total blindness users (average score)	Low vision users (average score)	Overall average score
Functionality and practicality	4.2	4.5	4.3
Ease of operation	3.9	4.1	4.0
Accuracy of information	4.3	4.4	4.3
Sense of security	4.4	4.2	4.3

As shown in *Table 4*, visually impaired users reported high overall satisfaction with the platform (average score 4.2), with the highest ratings for information accuracy and security perception, indicating recognition of the detection reliability demonstrated by the I-YOLOv8-BVOD model. Completely blind users rated

operational convenience low (3.9), primarily citing the need to simplify voice command logic; low-vision users suggested adding screen brightness auto-adjustment functionality. Interviews revealed that 80% of users believed the platform "significantly reduced travel anxiety", while 13% noted information overload in voice prompts for complex intersections. Future improvements will optimise priority-level sequencing for voice announcements, streamline content according to "distance-risk level-category" logic, and introduce customisable announcement frequency settings to enhance user experience.

4. DISCUSSION

The intelligent obstacle avoidance method for blind people based on improved YOLOv8 and binocular vision proposed in this study effectively solves the problems of poor adaptability to multiple scenarios and insufficient detection accuracy in traditional obstacle avoidance technology through algorithm optimisation and platform construction. The experimental results showed that the accuracy, recall and F1 score of the I-YOLOv8-BBOD model exceeded 0.90. In the confusion matrix analysis, false positives (FP=100) and false negatives (FN=50) were significantly lower than RetinaNet, DETR and traditional YOLOv8, demonstrating its advantages in reducing false positives and false negatives. The ablation experiment showed that the combination improvement of SPD Conv, C2f PC and SIOU loss functions increased mAP and AP50 by 13.19% and 16.00%, respectively. In the on-site testing of the platform, the successful obstacle avoidance rate exceeded 90%, the false alarm rate was less than 1.3%, and visually impaired users scored 4.3 points for information accuracy and security, verifying the practicality of the technology.

Compared with the intelligent head-mounted obstacle avoidance device proposed by Xu et al. [22], this study has more advantages in detection robustness and scene adaptation. Xu et al. implemented wearable integration through model screening, focusing on device portability. The research improved the structure of SPD Conv and PConv, which not only reduced computational costs to adapt to mobile devices but also optimised feature extraction for extreme scenarios such as strong light and rainy nights, resulting in an average increase of 12.8% in mAP compared to traditional YOLOv8, solving the problem of head-mounted devices being susceptible to environmental light interference. Compared with the indoor blind guidance study by Lv et al. [23], this study focuses on complex outdoor scenes and covers 15 types of obstacles and 5 types of road condition judgment objects through the WOTR dataset, which is more closely related to the actual travel needs of visually impaired people.

In summary, this study has made breakthroughs in algorithm improvement, scene coverage and system integration, but performance still declines in extreme scenarios. In the future, Xu et al.'s wearable device design experience can be combined to further optimise hardware integration. At the same time, drawing on the trajectory prediction technology of Lv et al., the entire process of "detection obstacle avoidance navigation" is optimised.

5. CONCLUSION

With the growth of DL technology, traditional obstacle avoidance methods for blind people can no longer meet practical needs. In response to the problems of poor adaptability to multiple scenarios and low obstacle avoidance rates in current obstacle avoidance methods for blind people, this study proposed an I-YOLOv8-BBOD model. The model first improved the YOLOv8 algorithm by introducing the SIOU function, SPD-Conv and a more efficient PConv structure, and then built BTIOAP based on the improved YOLO algorithm. In confusion matrix analysis, I-YOLOv8-BBOD performed the best and had a significant advantage in reducing FP and FN. Its precision was 0.90, its recall rate was 0.95, and its F1 score was 0.92, indicating the best overall performance. In the ablation experiment, the introduction of the SPD-Conv structure increased mAP and AP50 by 6.25% and 6.18%. After using C2f-PC, mAP and AP50 increased by 1.39% and 1.45%. After improving the loss function, mAP and AP50 increased by 2.08% and 1.82%. After the improvement of each module, mAP and AP50 increased by 13.19% and 16.00%, significantly improving detection accuracy. In the instance application, the successful obstacle avoidance rate of BTIOAP exceeded 90%, while the highest on traditional platforms was only 87.4%. However, the FAR of this platform was below 1.3%, and the highest on traditional platforms was 5.4%. This indicates that the platform can accurately detect obstacles and road conditions, providing safe and efficient travel solutions for visually impaired groups. However, the proposed method does not deeply analyse the different needs and adaptability of different types of blind users (such as different degrees of visual impairment) when using the platform. To address the diverse needs of users with various

types and degrees of visual impairment, we will conduct stratified adaptation research in the future. The project plans to collect data from 120 visually impaired individuals through questionnaires and in-depth interviews (including 40 fully blind users, 50 low-vision users and 30 users with combined physical disabilities), establishing user demand profiles. For fully blind users, we will enhance voice interaction by streamlining commands and improving feedback timeliness, keeping single-command length under 5 characters and response latency below 0.5 seconds. Low-vision users will benefit from a new high-contrast screen contour annotation feature with customisable colour and line width settings. Combined physical-impaired users will access a voice-activated contactless operation mode, enabling emergency obstacle avoidance triggered by "two consecutive wake-up commands". Additionally, we will integrate a user type selection module into the platform to automatically match functional configurations based on input, thereby enhancing system adaptability.

ACKNOWLEDGEMENTS

This work was supported by the Major Natural Science Research Project of Anhui Universities (KJ2021ZD0174) and the Visiting and Training Program for Outstanding Young Backbone Teachers of Anhui Universities (gxgnfx2021181).

REFERENCES

[1] Badrloo S, et al. Image-based obstacle detection methods for the safe navigation of unmanned vehicles: A review. *Remote Sensing*. 2022;14(15):3824-3839. DOI: 10.3390/rs14153824

[2] Alshammrei S, et al. Improved Dijkstra algorithm for mobile robot path planning and obstacle avoidance. *CMC-Computers Materials & Continua*. 2022;72(3):5939-5954. DOI: 10.32604/cmc.2022.024139

[3] Zhao M, et al. Single-frame infrared small-target detection: A survey. IEEE Geoscience and Remote *Sensing Magazine*. 2022;10(2):87-119. DOI: 10.1109/MGRS.2021.3122241

[4] Lei F, et al. Underwater target detection algorithm based on improved YOLOv5. J*ournal of Marine Science and Engineering*. 2022;10(3):310-322. DOI: 10.3390/jmse10030310

[5] Li H, et al. Rethinking pseudo labels for semi-supervised object detection. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022;36(2):1314-1322. DOI: 10.1609/aaai.v36i2.20164

[6] Yang J, et al. St3d++: Denoised self-training for unsupervised domain adaptation on 3d object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022;45(5):6354-6371. DOI: 10.1109/TPAMI.2022.3155392

[7] Rui C, et al. A multi-sensory blind guidance system based on YOLO and ORB-SLAM. *2021 IEEE International Conference on Progress in Informatics and Computing (PIC). IEEE*. 2021;1(1):409-414. DOI: 10.1109/PIC52629.2021.9629253

[8] Mu H, et al. Dynamic obstacle avoidance system based on rapid instance segmentation network. *IEEE Transactions on Intelligent Transportation Systems*. 2023;25(5):4578-4592. DOI: 10.1109/TITS.2022.3226003

[9] Cao X, et al. Research on obstacle detection and avoidance of autonomous underwater vehicle based on forward-looking sonar. *IEEE Transactions on Neural Networks and Learning Systems*. 2022;34(11):9198-9208. DOI: 10.1109/TNNLS.2022.3156907

[10] Rahman S, et al. Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography. J*ournal of Imaging*. 2023;9(10):216-230. DOI: 10.3390/jimaging9100216

[11] Wang D, et al. Farmland obstacle detection from the perspective of uavs based on non-local deformable detr. *Agriculture*. 2022;12(12):1983-1995. DOI: 10.3390/agriculture12121983

[12] Xue J, et al. Detection of farmland obstacles based on an improved YOLOv5s algorithm by using CIoU and anchor box scale clustering. *Sensors*. 2022;22(5):1790-1804. DOI: 10.3390/s22051790

[13] Xu Z, et al. Onboard dynamic-object detection and tracking for autonomous robot navigation with rgb-d camera. *IEEE Robotics and Automation Letters*. 2023;9(1): 651-658. DOI: 10.1109/LRA.2022.3228434

[14] Suman S, et al. Vision navigator: a smart and intelligent obstacle recognition model for visually impaired users. *Mobile Information Systems*. 2022;2022(1):9715891-9715900. DOI: 10.1155/2022/9715891

[15] Balasundaram A, et al. Zero-DCE++Inspired Object Detection in Less Illuminated Environment Using Improved YOLOv5. *Computers, Materials & Continua*. 2023;77(12):2751-2769. DOI: 10.32604/cmc.2023.039437

[16] Li M, et al. A method for top view pedestrian flow detection based on small target tracking. *Informatica*. 2024;48(11):1813.-1830. DOI: 10.15388/Informatica.2024.48.1813

[17] Liu K, et al. Underwater target detection based on improved YOLOv7. *Journal of Marine Science and Engineering*. 2023;11(3):677-690. DOI: 10.3390/jmse11030677

[18] Yang R, et al. KPE-YOLOv5: an improved small target detection algorithm based on YOLOv5. *Electronics*. 2023;12(4):817-830. DOI: 10.3390/electronics12040817

[19] Xie Z, et al. Optimized Method for Basketball Game Judging by Integrating Faster-RCNN with LK Algorithm. *Informatica*. 2024;48(23):17-31: 3367-3372. DOI: 10.15388/Informatica.2024.48.3367

[20] Chen G, et al. YOLOv8-CML: A lightweight target detection method for Color-changing melon ripening in intelligent agriculture. *Scientific Reports*. 2024;14(1):14400-14414. DOI: 10.1038/s41598-024-64824-4

[21] Hasanvand M, et al. Machine learning methodology for identifying vehicles using image processing. *Artificial Intelligence and Applications*. 2023;1(3):170-178. DOI: 10.54254/2755-2721/1/3/100020

[22] Xu P, et al. Intelligent head-mounted obstacle avoidance wearable for the blind and visually impaired. *Sensors*. 2023;23(23):9598-9611. DOI: 10.3390/s23239598

[23] Lv Z, et al. Blind travel prediction based on obstacle avoidance in indoor scene. *Wireless Communications and Mobile Computing*. 2021;2021(1):5536386-5536398. DOI: 10.1155/2021/5536386