Research on Object Detection Based on YOLO and Its Application in Tactile Paving Recognition

Zhixing Zhang

Beijing 21st Century School, Beijing, China

Abstract. The YOLO model, as an advanced real-time object detection algorithm, has made significant progress in the field of object detection since its inception. It achieves fast and accurate object detection by transforming the object detection task into a regression problem and using a single neural network to predict multiple bounding boxes and category probabilities. The YOLO model has significant advantages in processing speed, detection accuracy, and real-time performance, and is widely used in fields such as autonomous driving, intelligent monitoring, and medical image analysis. This paper adopts a tactile paving detection method based on YOLOv8 and CBAM attention mechanism, and improves the model's ability to extract tactile paving features by introducing CBAM attention mechanism and loss function. This method shows good adaptability when dealing with the tactile paving environment under different lighting and weather conditions, can effectively identify diverse obstacles, and provides a strong guarantee for the travel safety of the visually impaired. In conclusion, YOLO model has broad prospects in object detection and tactile paving detection. The tactile paving detection method based on YOLOv8 and CBAM attention mechanism adopted in this paper provides new ideas and methods for the development of tactile paving detection technology, and has important practical application value. In the future, with the continuous optimization of performance and technological innovation of YOLO series models, further improvement of hardware performance, and the fusion of YOLO and other sensor data to form a multimodal detection system, YOLO models are expected to play an important role in more practical application scenarios.

Keywords: YOLO Model; Tactile Paving Recognition; Object Detection.

1. Introduction

1.1 Research Background

Object detection is an important research task in computer vision, aimed at accurately discovering and locating objects in images or videos through automated methods, and assigning corresponding category labels to each detected object. Traditional object detection algorithms typically include steps such as candidate region selection, feature extraction, object classification and detection, and post-processing. However, under complex road conditions, these methods have an increased rate of missed and false detections due to the overlapping and occlusion of dense objects.

In order to address the limitations of traditional object detection algorithms, Joseph Redmon et al. proposed the YOLOv1 detection algorithm in 2016. This model applies a single neural network to the entire image, significantly improving detection speed and enhancing localization accuracy for small-sized objects. Subsequently, YOLOv2 and YOLOv3 models were successively introduced, further improving accuracy, speed, and the number of detection categories. The core idea of the YOLO model is to treat object detection tasks as regression problems, directly mapping from image pixels to bounding box coordinates and category probabilities through a single neural network. This innovative method significantly improves detection speed and accuracy, making YOLO a leading technology in the field of object detection.

Visually impaired individuals are unable to perceive their surroundings due to visual impairments, resulting in numerous difficulties and inconveniences during their travels. Although the existing blind guidance technology has realized the functions of obstacle detection and tactile paving segmentation to a certain extent, it has problems such as low detection accuracy and incomplete feedback information. In this context, YOLO, as an advanced real-time object detection algorithm, has become an ideal choice for tactile paving detection because of its efficiency and accuracy. YOLO can quickly

identify and locate the tactile paving by monitoring the video stream of urban roads in real time, and find out the situation blocked or damaged by obstacles in time, so as to provide necessary data support for urban managers and ensure the effectiveness and safety of the tactile paving.

In addition to tactile paving detection, YOLO also plays an important role in other fields. In the field of autonomous driving, YOLO is capable of real-time detection of pedestrians, vehicles, and obstacles, enhancing the safety and responsiveness of vehicles. In intelligent monitoring systems, YOLO quickly identifies abnormal behavior and enhances public safety. In medical imaging analysis, YOLO is used to quickly detect lesion areas and assist doctors in early diagnosis.

1.2 Research Purpose and Significance

The research on object detection applications based on YOLO is of great significance in promoting technological progress in related fields and solving practical problems. The YOLO algorithm has become a leading technology in the field of object detection due to its high efficiency and real-time performance, and is widely used in multiple fields such as autonomous driving, intelligent monitoring, and medical image analysis.

The YOLO model achieves fast detection speed by transforming object detection tasks into regression problems and using a single neural network to predict multiple bounding box and category probabilities. Compared to other object detection algorithms, the YOLO model has significantly improved detection speed. The YOLO model improves the detection accuracy of objects of different scales and shapes by using techniques such as multi-scale feature extraction and anchor box mechanism. Especially when dealing with small and overlapping objects, the YOLO model exhibits better performance.

The design concept of YOLO model is to achieve real-time object detection; therefore, it has significant advantages in processing speed. This makes the YOLO model have broad application prospects in application scenarios that require rapid response, such as autonomous driving and intelligent monitoring. Its fast-processing capability makes real-time detection possible, effectively identifying objects in complex environments, greatly improving the system's response speed and accuracy.

In the context of tactile paving detection, tactile paving provides a safe path for visually impaired people, but obstacles, pedestrians and other traffic conditions in the urban environment often affect the effectiveness of tactile paving. Through the application of YOLO algorithm, the system can identify obstacles on the tactile paving in real time to ensure the safe travel of visually impaired people. The specific goal of this study is to optimize the application of YOLO algorithm in tactile paving detection, and improve its detection accuracy and speed in complex environments. The research will focus on several key aspects: adjust the parameters of YOLO model to improve its adaptability according to the tactile paving environment under different lighting and weather conditions; Utilize data augmentation techniques to expand the training dataset and enhance the model's ability to detect diverse obstacles; Combining deep learning and image processing technology, a complete tactile paving detection system is developed to realize real-time monitoring and feedback functions. This will not only enhance the travel safety of visually impaired individuals, but also provide real-time data support for urban managers to help them more effectively maintain and optimize accessible environments.

1.3 Conclusion

This chapter mainly discusses the background of object detection technology and its importance in modern society, especially in the field of tactile paving detection. With the rapid development of artificial intelligence and deep learning, the accuracy and efficiency of object detection technology have been significantly improved, but still face many challenges in complex environments, especially in tactile paving detection, how to effectively identify obstacles to ensure the safety of visually impaired people is particularly important.

2. Literature Review

2.1 Current Status of Object Detection Research

Traditional object detection algorithms typically include several steps: candidate region selection, feature extraction, object classification and detection, and post-processing. In obstacle object detection, commonly used feature extraction methods include Haar, Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and Speeded Up Robust Features Algorithm (SURF). For example, Wei et al. [1] used Haar and HOG features for feature extraction of target vehicles and classified them using Support Vector Machine (SVM). However, these methods extract limited shallow feature information, making it difficult to achieve accurate classification and precise localization. Especially under complex road conditions, the overlap and occlusion of dense objects lead to an increase in missed and false detection rates.

The emergence of convolutional neural networks has brought new ideas to the field of object detection. Girshick et al. proposed the RCNN model in 2014 [2]. The model first uses selective search techniques to select candidate boxes that may contain objects, then adjusts each candidate box to a fixed size, inputs it into a pre trained CNN model to extract features, and finally uses an SVM classifier to determine whether there are objects and recognize categories. Subsequently, models such as Fast RCNN [3] and Faster RCNN [4] were introduced successively. Although they have improved detection accuracy and speed, there is still computational redundancy. In 2016, Joseph Redmon et al. proposed the YOLOv1 detection algorithm [5]. This model applies a single neural network to the entire image, significantly improving detection speed, but with a slight decrease in localization accuracy for small-sized objects. In 2017, the team improved the accuracy, speed, and number of detection categories based on YOLOv1 and launched YOLOv2, known as YOLO9000 [6] due to its ability to detect 9000 objects. YOLOv2 uses a brand new fully convolutional feature extraction network Darknet-19 and introduces anchor box mechanism, significantly improving recall and accuracy. However, there is still room for improvement in detecting high overlap and small objects. YOLOv3 [7] adopts a more powerful feature extractor and enhances model performance through a multi-scale feature extraction architecture, enabling it to better detect objects at various scales. With the continuous development of the YOLO algorithm, speed and accuracy are gradually approaching a balance. Models such as YOLOv4 [8] and YOLOv5 have introduced new backbone feature extraction networks, data augmentation techniques, and optimization training strategies, significantly improving accuracy.

Nowadays, object detection has been widely applied in various fields. Some scholars have improved existing models for obstacle detection. For example, Cong and Li [9] used YOLOv3 for obstacle detection, and experimental results showed that compared with YOLOv2, YOLOv3 has better detection performance under different lighting conditions. However, YOLOv3 has poor handling of overlapping objects, and the samples are limited to large objects such as cars and people, resulting in low generalization performance. Benjumea et al. [10] optimized YOLOv5's feature extraction ability for small objects by redirecting feature maps to the neck network and head layer. Guan et al. [11] proposed a more accurate and high-speed model for railway obstacle detection tasks based on YOLOv4-tiny. The backbone network integrates the MOSA module with attention mechanism, and residual blocks are embedded in the neck and head to enhance feature extraction capability. He et al. [12] focused on Neural Architecture Search (NAS) technology, using NAS to search for feature pyramid networks and anchor free object detection models in the prediction stage. The average accuracy of YOLO-NAS was 2.6% higher than the original base model.

2.2 Current Status of Tactile Paving Detection Research

In recent years, many scholars at home and abroad have applied image processing technology to tactile paving detection. For example, Yuki et al. [13] proposed tactile paving detection based on color information of fixed threshold, and Yamanaka et al. [14] proposed a detection system for yellow tactile paving. However, under different lighting conditions or tactile paving color changes, the color

and saturation in the captured image may change significantly, so the effect of fixed threshold method in tactile paving detection is not ideal. On this basis, Ito et al. [15] extracts the HSV histogram within the boundary line of the tactile paving, statistically determines the dynamic threshold, and finally applies it to the tactile paving area detection, with improved detection accuracy.

With the rapid development of deep learning, it has been widely applied in the field of image segmentation. Unlike traditional image segmentation methods, deep learning does not require manual feature design. Aktaş et al. [16] fused the object detection model YOLOv3 with DenseNet for tactile paving detection, and marked the tactile paving area with the smallest rectangular box to feedback the category and location information. In contrast, Chen Qianqian et al. [17] used YOLOv5 model to improve feature extraction. After comparing models of different scales, they finally selected YOLOv5s as the detection model of tactile paving and zebra crossing. However, these object detection models can not accurately segment the tactile paving area, so some scholars have adopted semantic segmentation models. For example, Yang et al. [18] used real-time semantic segmentation to provide accessible regional awareness for visually impaired people, including tactile paving, sidewalks and stairs. Cao et al. [19] proposed a lightweight deep convolutional network based on MobileNet, while Xia et al. [20] replaced the encoder of SegNet with the feature extraction part of MobileNetV3 to improve detection speed. However, the lightweight model performed poorly in terms of detection accuracy.

2.3 Summary of this Chapter

Although significant progress has been made in assistive technology for the visually impaired, challenges still exist. For example, there are relatively few obstacle objects and tactile paving data sets for the travel of the visually impaired, which limits the performance of the deep learning model in specific scenarios. Lack of sufficient data for training makes it difficult for the model to learn enough features for accurate detection. Meanwhile, the detection range and accuracy of existing technologies still need to be improved to meet the requirements of complex environments.

3. Theoretical Basis and Key Technologies

3.1 Fundamentals of Deep Learning

Convolutional Neural Network (CNN) is a deep learning architecture specifically designed for processing and analyzing image data, widely used in tasks such as image classification, object detection, and semantic segmentation. Its core components include convolutional layers, pooling layers, fully connected layers, and activation functions. The core function of convolutional layers is to extract features of the input image through filters (convolution kernels). Its main feature is weight sharing, which means that the same convolution kernel slides across the entire input, significantly reducing the number of model parameters and lowering computational complexity. In addition, convolutional layers have local perceptual properties, allowing each output pixel to be associated only with the local region of the input, effectively extracting fine-grained features such as edges and textures. The role of the pooling layer is to reduce the spatial dimension of the feature map, decrease computational complexity, and preserve important features. The commonly used pooling methods include max pooling and average pooling. The former selects the maximum value of local regions to preserve salient features, while the latter suppresses noise and preserves details by calculating the average value of local regions. The fully connected layer integrates the previously extracted features and learns global features. Its structure is that each neuron is connected to all neurons in the previous layer, capturing complex relationships between features by adjusting weights and biases, and ultimately mapping them to the output space. The activation function provides nonlinear transformations in this process, enabling the network to learn complex nonlinear relationships and enhance the model's expressive power.

3.2 Object Detection Methods

Object detection is an important task in the field of computer vision, aimed at accurately discovering and locating objects in images or videos through automated methods, and assigning corresponding category labels to each detected object. At present, deep learning-based object detection algorithms are mainly divided into Two Stage methods and One Stage methods. The core concept of the Two Stage method involves two key steps, namely generating Region Proposals and classifying and finely locating these candidate regions, which has made significant progress in accuracy and localization precision. Relatively speaking, the One Stage method utilizes a single feedforward propagation process to simultaneously generate candidate regions, perform object classification, and refine localization, achieving more efficient and rapid object detection. The uniqueness of the One Stage method lies in transforming the object detection task into a regression problem, where YOLO (You Only Look Once) is considered a typical algorithm in the field of one-stage object detection.

3.3 Introduction to YOLO Algorithm

The YOLO (You Only Look Once) algorithm is an efficient object detection method, whose core idea is to treat object detection tasks as regression problems. This algorithm significantly improves detection speed and accuracy by directly mapping image pixels to bounding box coordinates and category probabilities through a single neural network. YOLO divides the entire image into SxS fixed size grid cells, with each cell predicting B bounding boxes. Each bounding box is defined by five parameters: center coordinates (x, y), width w, height h, and the confidence of the object's existence. The formula for calculating confidence is:

$$Confidence = P_r * IOU_{Truth}^{pred}$$

Among them, P_r is a binary label indicating whether the grid cell contains the object. If it does, it is set to 1; otherwise, it is set to 0. Meanwhile, IOU_{Truth}^{pred} is the intersection and union ratio between the predicted bounding box and the real bounding box.

3.4 Introduction to YOLO Network Structure

The algorithm process of YOLO is implemented through Convolutional Neural Networks (CNN). After multiple convolutional and pooling layers are applied to the input image, a feature map is output, with each cell of the feature map corresponding to a grid. To eliminate overlapping bounding boxes, YOLO adopts Non-Maximum Suppression (NMS) technique. The architecture of YOLOv8 mainly consists of three parts: Backbone, Neck, and Head. The YOLOv8 network architecture diagram is shown in Figure 1

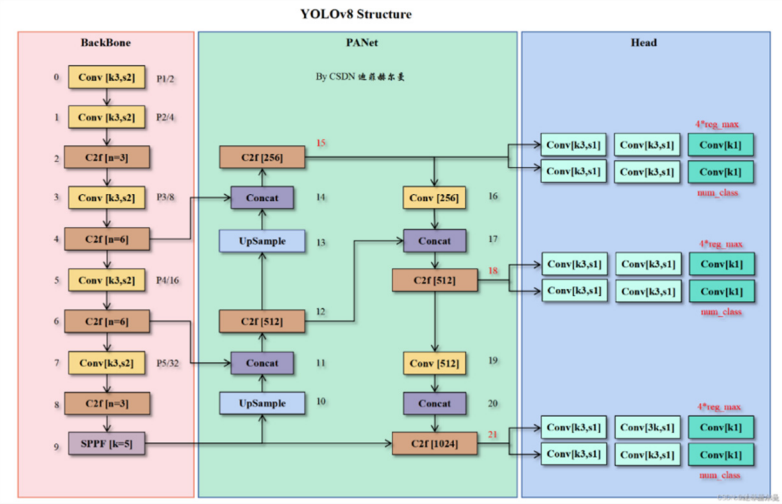


Figure 1. YOLOv8 Network Structure Diagram

(1) Backbone. Backbone is the feature extraction part of YOLOv8, responsible for extracting effective features from input images. It processes images through multiple convolutional layers and activation functions to generate feature maps at different levels. YOLOv8 typically uses efficient network structures such as CSPNet (Cross Stage Partial Network) to improve the speed and accuracy of feature extraction. Ultimately, Backbone will output multiple feature layers of different shapes, which contain information of different scales and semantics in the image.

(2) Neck. The main function of the Neck section is to fuse different scale features from Backbone. Through techniques such as Feature Pyramid Network (FPN) or Path Aggregation Network (PANet), Neck can effectively combine high-level and low-level feature information, thereby enhancing the model's ability to detect objects of different scales. This feature fusion helps improve the performance of the model in complex scenes, especially in the detection of small or overlapping objects.

(3) Head. Head is the output part of YOLOv8, responsible for predicting the position, category, and confidence of the object. By processing the fused features output by Neck, the Head generates the final detection result, including the position coordinates of the bounding box, the object category, and its confidence score. YOLOv8 adopts an efficient output mechanism to improve detection speed and accuracy, ensuring fast response in real-time applications.

4. Tactile Paving Detection Based on YOLO

This paper proposes a YOLOv8 object detection model based on attention mechanism. By introducing attention mechanism, the task of tactile paving segmentation and detection is completed at the same time.

4.1 Dataset Construction

The blind guidance system collects pictures of the tactile paving of visually impaired people when they walk through the front camera. However, due to the influence of light, the images are easy to blur, details are lost, and noise may also be generated in the transmission process. These issues will lower the image quality, thereby affecting the subsequent image segmentation results. Therefore, image preprocessing is necessary to improve quality, highlight object features, and reduce interference information. Common preprocessing methods include image enhancement and morphological processing.

4.1.1 Image Preprocessing

Due to the wide range of image sources, there is a problem of inconsistent image quality. For example, in low light conditions, the object in the image is not clear enough to effectively identify the relevant information of the object, and it is necessary to enhance the image to highlight the object area. Image enhancement technology can enhance the contour information of the object and reduce the impact of noise on subsequent processing. The image enhancement method in this article achieves the enhancement of contrast and clarity of images through color space conversion and histogram equalization. Convert an RGB image to the YCbCr color space to separate luminance information (Y) and chrominance information (Cb and Cr), with the conversion formula including the calculation of Y as:

$$Y = 0.299R + 0.587G + 0.114B,$$

$$Cb= -0.1687R - 0.3313G + 0.5B + 128$$

$$Cr= 0.5R - 0.4187G - 0.0813B + 128$$

Next, extract the Y component for histogram equalization, count the histogram of the Y component, calculate the number of pixels for each gray level, and form a histogram distribution map. Then, based on the distribution of the histogram, adjust the grayscale values of the Y component to make their

distribution more uniform. Use the following formula to calculate the number of pixels for each grayscale value:

$$P(r_k) = \frac{n_k}{W * H}$$

Where W and H are the width and height of the image. After processing, update the Y component and combine it with the original Cb and Cr components. Finally, convert it back to the RGB color space to generate an image with higher contrast and clearer details, thereby improving the overall image quality.

The original image and the enhanced result are shown in Figure 2.

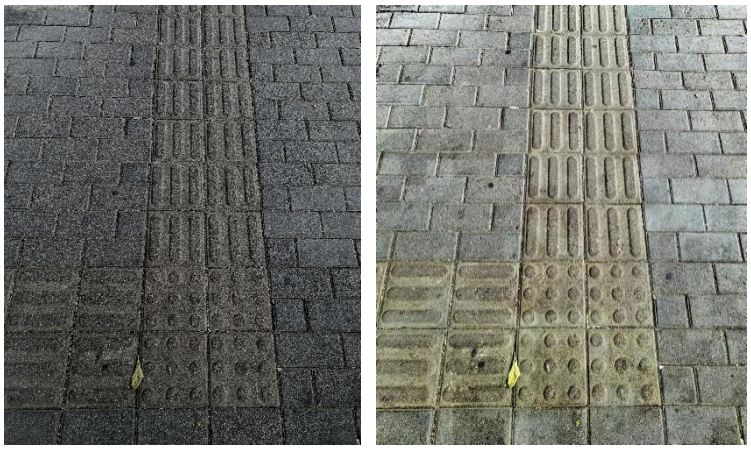


Figure 2. Image Enhancement Process

4.1.2 Image Annotation

The tactile paving data is 300 pictures including the tactile paving taken on the spot, which are manually marked with Labelme. When marking, the tactile paving area is marked. After selecting the object, save the results, get the corresponding JSON file, and run the script file in python to transfer the JSON file to the binary mask image. The annotation result is shown in Figure 3 below.

4.1.3 Image Enhancement

Due to the diverse input forms of data for the model and the overfitting caused by the small number of datasets, image enhancement is also required. By performing random cropping, random noise, random rotation, random brightness, random flipping, and other operations on the images, the dataset can be expanded to improve the model's normalization ability and robustness. The schematic diagram of data augmentation is shown in Figure 4:

4.2 Improve YOLOv8 Object Detection Algorithm

4.2.1 CBAM Attention Mechanism

CBAM (Convolutional Block Attention Module) is an attention mechanism module used to improve the performance of convolutional neural networks, combining channel attention and spatial attention mechanisms. The channel attention module aims to identify important channels, calculate the attention weights of channels through global average pooling and max pooling, and use these weights to recalibrate the channel features of the input feature map. The spatial attention module focuses on important spatial positions in the feature map. It calculates spatial attention weights by applying convolution operations in the spatial dimension and recalibrates the spatial features of the input feature map accordingly. The core idea of CBAM is to refine it in two stages: first, determine which channels are important, and then identify the spatial positions with information in the feature map. This dual attention mechanism enables CBAM to comprehensively capture key information in features, thereby improving the model's performance in various visual tasks.

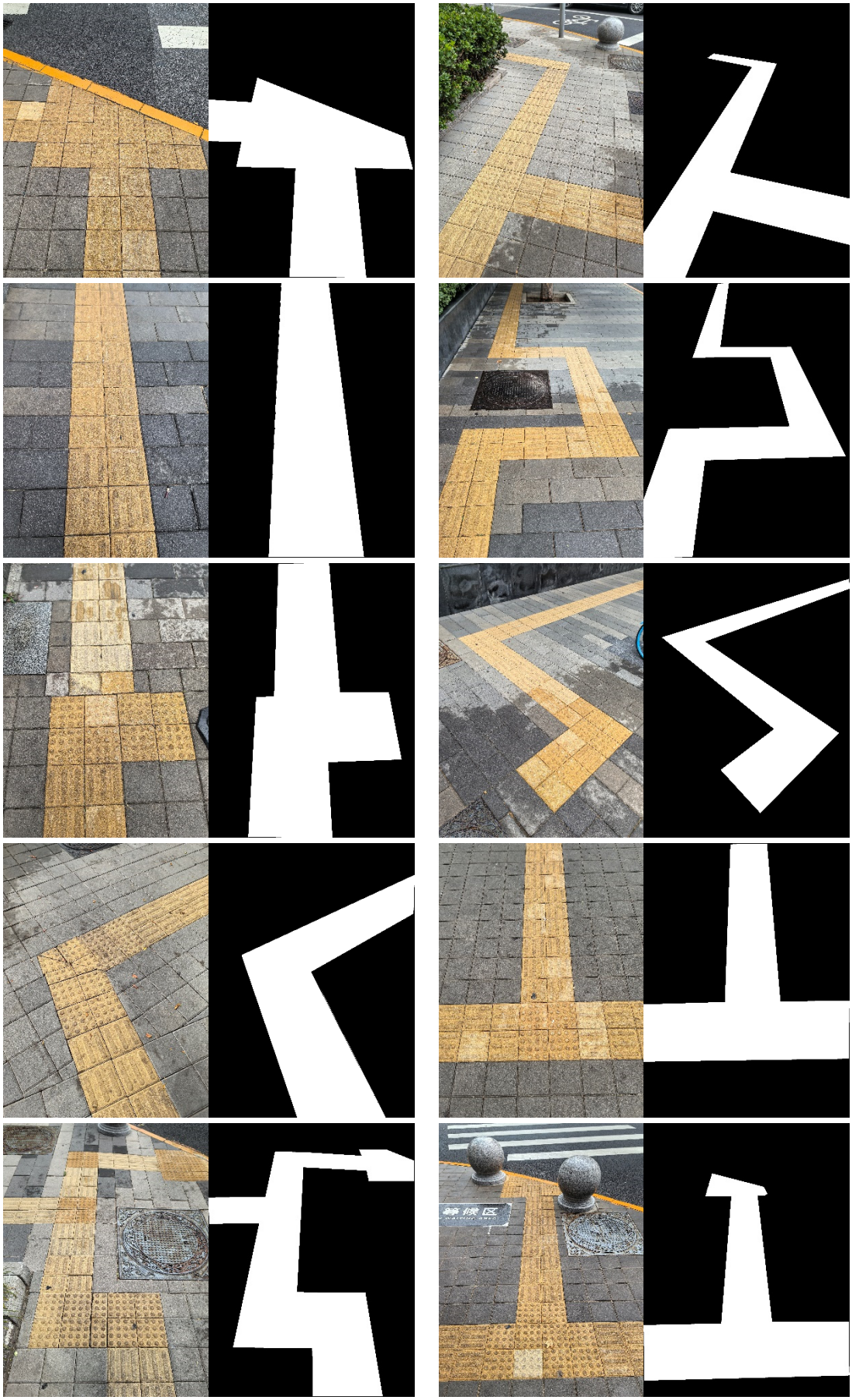


Figure 3. Partially Annotated Data

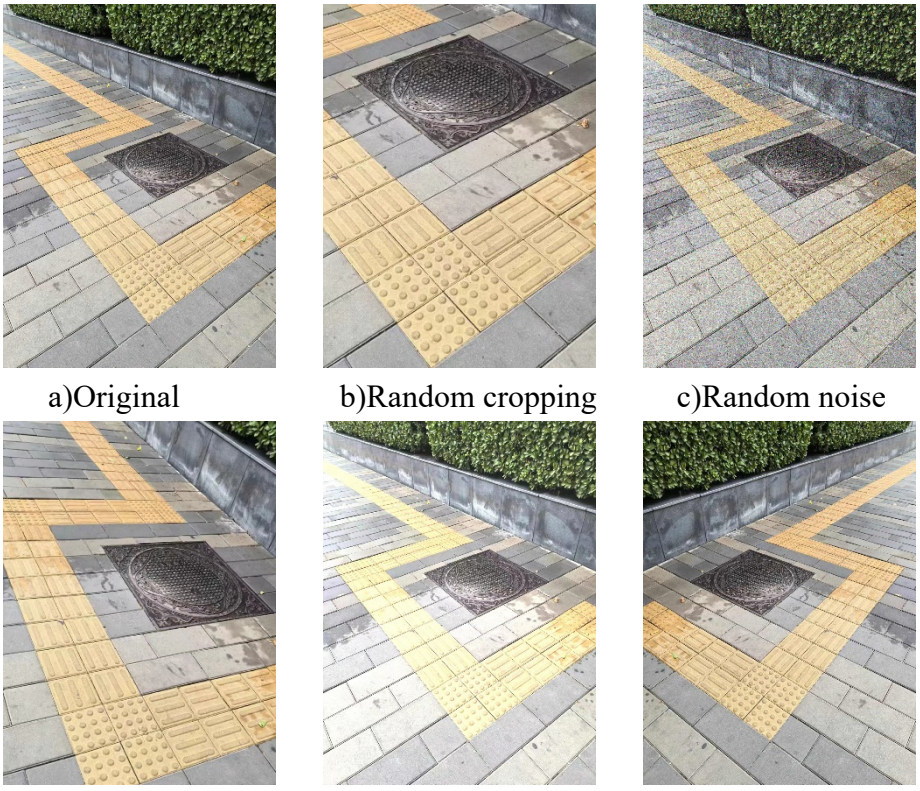


Figure 4. Schematic diagram of data augmentation

The goal of tactile paving detection is to accurately identify and locate the tactile paving signs on the road, and the introduction of CBAM module can improve the detection accuracy. By inserting the CBAM module into the feature extraction network of YOLOv8, the model can adaptively adjust the weight of the feature map to highlight the key features of the tactile paving. The CBAM module adopts residual connection method. The addition of the feature map processed by the attention mechanism and the original feature map can not only effectively alleviate the gradient disappearance problem in the deep network, but also retain the original features of the tactile paving. The weighted fusion strategy of CBAM module enables the model to give higher weight to the key areas of the tactile paving. This adaptive feature fusion strategy is particularly important when dealing with tactile paving detection, because the shape and color of tactile paving are often significantly different from the surrounding environment.

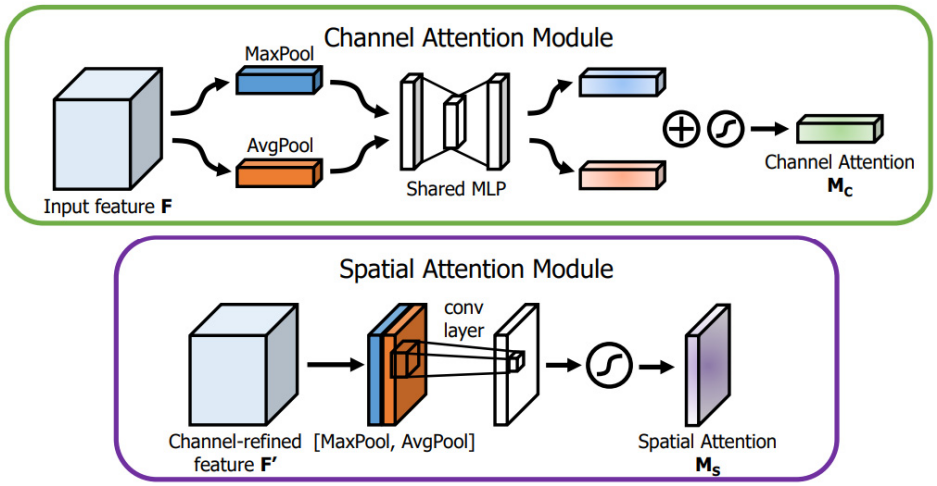


Figure 5. CBAM Attention Mechanism Structure Diagram

The CBAM module consists of two parts: channel attention and spatial attention. The channel attention part generates channel descriptors through global average pooling and max pooling, and then generates channel weights through a shared fully connected layer. Finally, the weights are normalized using a sigmoid activation function and applied to the input feature map to achieve channel feature recalibration. Next, the spatial attention part concatenates the feature maps processed by channel attention on the channel dimension and generates spatial attention maps using convolution operations. The sigmoid activation function is also used to normalize the weights and apply them to the feature maps output by channel attention to achieve spatial feature recalibration. The structure diagram of CBAM attention mechanism is shown in Figure 5:

4.2.2 Loss Function

In the task of tactile paving detection, the object is to accurately identify and locate the tactile paving area. For this purpose, we have designed a comprehensive loss function that includes the following parts:

(1) Cross Entropy Loss. In order to improve the model's ability to distinguish between objects and backgrounds, this paper uses cross entropy loss to optimize confidence prediction. Cross entropy loss can effectively handle binary classification problems, ensuring that the model can accurately determine whether each box contains the object. The following equation shows the disclosure of cross entropy loss:

$$L_{conf} = -\frac{1}{P}\sum_{i=1}^{p}[y_i * \ln \hat{y}_i + (1 - y_i) * \ln(1 - \hat{y}_i)]$$

Among them, P is all pixels in the image, $y_i$is the true label, and$\hat{y}_i$ is the predicted value.

(2) Focal loss. Considering the class imbalance in tactile paving detection, Focal Loss is introduced to optimize the classification loss. Focal Loss adjusts sample weights to make the model focus more on difficult to classify samples during training. The following formula shows the Focal Loss announcement:

$$L_{conf} = -\frac{1}{N}\sum_{i=1}^{N}\alpha(1 - P_i^{pred})^{\gamma}\ln(P_i^{pred})$$

Among them, α is the balance factor, and γ is the focusing factor.

5. Experimental Results and Analysis

5.1 Experimental Environment and Training Process

To verify the effectiveness of the object detection model proposed in this article, experiments were conducted on a self-made dataset. The computer processor model used in the experiment is R5-5600U, the graphics card model is MX450 GDDR6, and the operating system is Windows 10.

5.2 Training Parameter Settings

During the training process, Epoch is set to 100, Batch size is set to 2, and the learning rate is set to 0.001, in conjunction with the Adam optimizer for training. Adopting the early stopping method with decreasing learning rate for training, if the loss value of the validation set does not decrease three times, the learning rate will decrease, and then continue training. When the loss value of the validation set does not decrease continuously, it means that the model training is basically completed and the training will be stopped. During the experiment, training was stopped after 64 iterations, and the accuracy and loss curves of the training and validation sets are shown in Figure 6. The horizontal axis represents the number of iterations, while the vertical axis represents the accuracy and loss values, respectively.

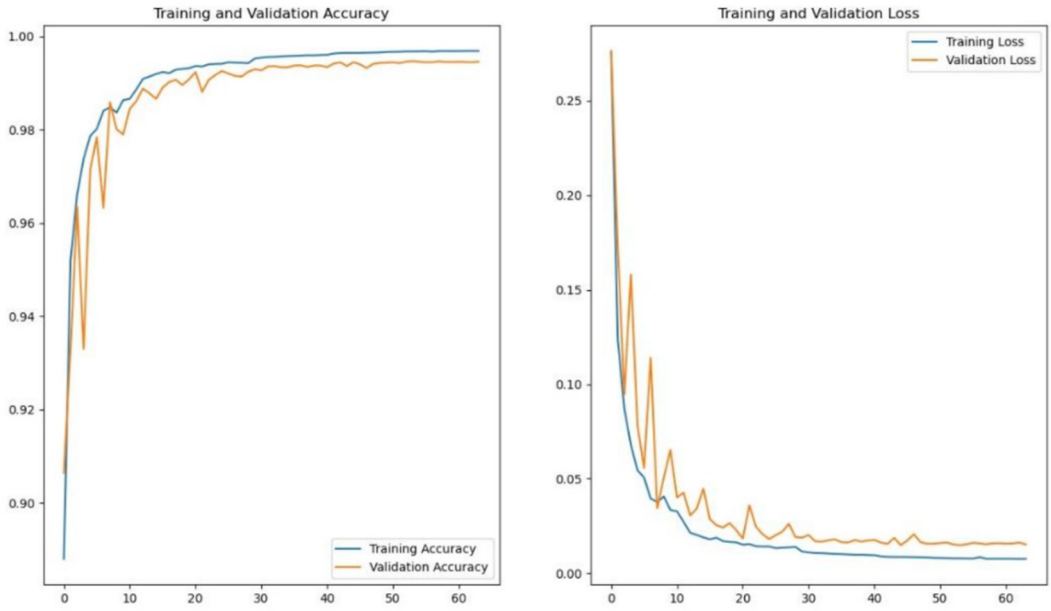


Figure 6. Accuracy curve and Loss curve during the experimental process

5.3 Analysis of Experimental Results

In order to verify the effectiveness of the model in this article, after training the model, it was tested on the test set, and the partial test results are shown in Figure 7. As shown in the figure, the object detection system based on YOLOv8 proposed in this paper can effectively segment the tactile paving with strong adaptability. It can segment the tactile paving with obvious color features, such as yellow tactile paving, dark gray tactile paving, etc. Even tactile paving images with insignificant color features and texture features can be accurately identified.



Figure 7. Object Detection Test Results

6. Discussion and Prospect

6.1 Summary of Research Findings

The main finding of this study is that the accuracy and speed of detection in complex environments can be improved by optimizing the application of YOLO algorithm in tactile paving detection. In this paper, the parameters of YOLO model are adjusted according to the tactile paving environment under

different lighting and weather conditions to improve its adaptability. Meanwhile, we utilized data augmentation techniques to expand the training dataset and enhance the model's ability to detect diverse obstacles. In addition, we have also developed a complete tactile paving detection system based on deep learning and image processing technology, realizing real-time monitoring and feedback functions. The innovation of this study is mainly reflected in the introduction of CBAM attention mechanism. By introducing CBAM (attention mechanism) into YOLOv8 model, the ability of the model to extract tactile paving features is improved.

6.2 Limitations and Challenges

Although this study has achieved certain results in the field of YOLO based object detection applications, there are still some limitations and challenges. First of all, there are relatively few obstacle objects and tactile paving data sets for the travel of the visually impaired, which limits the performance of the deep learning model in specific scenarios, and makes it difficult for the model to learn enough features to accurately identify. Secondly, the detection range and accuracy of existing technologies still need to be improved to meet the requirements of complex environments, especially when dealing with small or overlapping objects, the generalization ability of the model needs to be improved.

6.3 Future Research Directions

The future research direction can be further refined and expanded. For example, consider cooperating with the city management department to expand the data set, collect more tactile paving and obstacle data in different cities and road environments, so as to improve the applicability of the model in different regions. Meanwhile, lightweight network structures can be explored to optimize the model and improve its deployment efficiency on mobile devices. In addition, we can also try to combine tactile paving detection with other computer vision tasks such as semantic segmentation, instance segmentation, etc., and achieve richer scene understanding ability through multi task learning. The in-depth exploration of these research directions will help promote the development of YOLO based object detection technology, and make it play a greater role in practical applications.

7. Conclusion

This research has made significant contributions in the application field of object detection based on YOLO, mainly reflected in improving the application effect of YOLO algorithm in tactile paving detection. By optimizing model parameters, introducing CBAM attention mechanism, and designing a comprehensive loss function, the detection accuracy and speed of the model in complex environments have been improved. In addition, the research also developed a complete set of tactile paving detection system, realizing real-time monitoring and feedback functions, providing strong support for the travel safety of the visually impaired and the decision-making of city managers. The application value of research results mainly lies in improving the travel safety of visually impaired people, providing decision support for urban managers, and promoting technological development in related fields, providing new technological means for solving practical problems.

References

[1] Wei Y, Tian Q, Guo J, et al. multi-vehicle detection algorithm through combining Harr and HOG features [J]. Mathematics and Computers in Simulation, 2019, 155:130-145.

[2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[3] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[4] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[5] 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. 2016: 779-788.

[6] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[7] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprintarXiv: 180402767, 2018.

[8] Bochkovskiy A, Wang C-Y, Liao H-Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:200410934, 2020.

[9] Cong Z, Li X. Track obstacle detection algorithm based on YOLOv3[C]. 202013th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2020: 12-17.

[10] Benjumea A, Teeti I, Cuzzolin F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles[J]. arXiv preprint arXiv:211211798, 2021.

[11] Guan L, Jia L, Xie Z, et al. A lightweight framework for obstacle detection in the railway image based on fast region proposal and improved YOLO-tiny network[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-16.

[12] He C, Saha P. Investigating YOLO Models Towards Outdoor Obstacle Detection for Visually Impaired People[J]. arXiv preprint arXiv:231207571, 2023.

[13] Yuki I, Premachandra C, Sumathipala S, et al. HSV conversion based tactile paving detection for developing walking support system to visually handicapped people[C]. 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT). IEEE, 2019: 138-142.

[14] Yamanaka Y, Takaya E, Kurihara S. Tactile Tile Detection Integrated with Ground Detection using an RGB-Depth Sensor[C]. ICAART (2). 2020: 750-757.

[15] Ito Y, Premachandra C, Sumathipala S, et al. Tactile paving detection by dynamic thresholding based on HSV space analysis for developing a walking support system[J]. IEEE Access, 2021, 9: 20358-20367.

[16] Aktaş A, Doğan B, Demir Ö. Tactile paving surface detection with deep learning methods[J]. Journal of the Faculty of Engineering and Architecture of Gazi University, 2020, 35(3): 1685-700.

[17] Chen Qianqian, Wang Huan, Zhu Min Tactile paving and zebra crossing detection based on YOLOv5 [J] Information Technology and Informatization, 2022, (07): 10-14.

[18] Yang K, Bergasa L M, Romera E, et al. Semantic perception of curbs beyond travers ability for real-world navigation assistance systems[C]. 2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES). IEEE,2018: 1-7.

[19] Cao Z, Xu X, Hu B, et al. Rapid detection of blind roads and crosswalks by using a light weight semantic segmentation network[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(10): 6188-6197.

[20] Xia Y, Li Y, Ye Q, et al. Image segmentation for blind lanes based on improved SegNet model[J]. Journal of Electronic Imaging, 2023, 32(1): 013038-013038.