DOI:10.18686/ahe.v7i35.12523

Smart Guide Glasses Based on YOLOv5 and Binocular Ranging Algorithms

Maolin Ye , Shaogeng Zeng *, Jinhao Qiu , Zhan Huang , Yuhang Deng

Lingnan Normal University School of Computer and Intelligent Education, Guangdong Zhanjiang.524048

Abstract: With the continuous development of science and technology, intelligent guide technology has become an important field to improve the quality of life of visually impaired people. YOLOv5 and binocular ranging technology, as key innovations in the field of computer vision and deep learning, bring new opportunities and challenges to intelligent guide blindness systems. This paper aims to deeply study the principle of YOLOv5 and binocular ranging technology, as well as their specific applications in the field of blindness guidance, focusing on their role and advantages in obstacle detection, distance measurement and navigation accuracy improvement.

Keywords: Intelligent guide; YOLOv5; Binocular ranging; Object detection

Fund Project:

Project fund: 2023 Guangdong College Student' innovation and entrepreneurship training program in 2023, Project No: Provincial Project No.:202310579001; 2021 Lingnan Normal University Students' Social Practice Teaching Base Project Construction Project, Project Name: College students ' social practice teaching base for artificial intelligence education

1. Research background and significance

With the rapid development of the field of computer vision and deep learning, YOLOv5 has become an efficient object detection tool, and its fast and accurate characteristics make it an ideal choice for intelligent blind guide systems. At the same time, binocular ranging technology can provide accurate distance information, which is expected to solve the problems that visually impaired people often face in navigation, such as obstacle detection and distance measurement.

2. System overview

Based on the embedded development module Raspberry Pi 4, combined with the YOLOv5 target detection algorithm and binocular ranging algorithm, we designed and developed a lightweight and versatile smart guide glasses. The product can detect obstacles and calculate the distance between the visually impaired and the obstacles, and broadcast to the visually impaired by using the offline voice module. This product is also equipped with SIM800C SMS communication module, which can send SMS to call for help to the contact person set in advance by pressing the button, and prevent the visually impaired from getting lost. In addition, this product is also equipped with a gyroscope module, which detects changes when a visually impaired person falls and sends a voice call for help to the surroundings under the control of the main control panel.

3. YOLOv5 algorithm and binocular ranging principle

3.1 Structure of YOLOv5

Input: Responsible for accepting the original image dataset and processing it into a feature representation suitable for model processing. The input side normalizes and normalizes the dimensions of the input image to ensure that the image is entered into the model at a uniform scale. In order to enhance the generalization ability and robustness of the model, YOLOv5 adopts the Mosaic method at the input end for data enhancement to provide richer training samples.

Backbone: It consists of multiple convolutional layers and pooling layers, which is the core of the entire model. The Backbone

network first performs feature extraction, gradually reducing the spatial size of the feature map through a series of convolutional layers while increasing the depth of the feature. This helps the model capture features at different levels, from edges and textures to more abstract semantic information. The pooling layer is used to reduce the spatial size of the feature map, which helps reduce the computation effort and improve the efficiency of the model.

Neck network: After the Backbone network, YOLOv5 adopts the neck network for feature reuse and multi-scale feature fusion. The network structure can effectively fuse feature maps of different scales to obtain richer feature representations.

Head network: contains three scales of predictive feature layers, corresponding to the detection of targets of different sizes. These prediction feature layers perform object detection according to feature maps of different scales, regression prediction of the detection frame and output the final detection result.

3.2 Activation function in YOLOv5 algorithm model structure

Activation functions play a crucial role in neural networks, introducing nonlinear properties that allow neural networks to capture complex patterns and features.

Mish (Mishra Activation Function) is a new type of activation function, which has soft saturation and is better able to handle the gradient disappearance problem.

$$\text{Mish}(x) = x \cdot \tanh(\text{softplus}(x))$$

The softplus function is defined as:

$$\text{softplus}(x) = \log(1 + e^x)$$

SiLU (Sigmoid Linear Unit) is a variant of the Sigmoid activation function. The SiLU function has a smooth nonlinear property, which contributes to the stable propagation of the gradient and accelerates the convergence rate of the model [1].

The mathematical definition of the SiLU function is:

$$\text{SiLU}(x) = x \cdot \sigma(x)$$

Where σ(x) is the Sigmoid function.

The selection of these activation functions is to enable YOLOv5 to better capture the target information in the image, so as to achieve efficient and accurate target detection.

3.3 Outputs in the YOLOv5 model structure

The output structure of YOLOv5 is an extension of its target detection backbone, which is responsible for converting the feature map into the final detection result. Anchor boxes are a key component of the YOLOv5 output. These predefined bounding boxes have different sizes and aspect ratios and are used to detect objects of different sizes and shapes. The model will use the information from these anchor boxes to predict the location and category of the target.YOLOv5 uses convolutional layers to process feature maps and generate predictions for target detection. These convolutional layers are responsible for extracting features and adjusting the parameters of the anchor frame to adapt it to the position and shape of the target. The output generates an "objectness score" that indicates whether each bounding box contains the target object. This score helps the model determine which bounding boxes actually contain the target, thus improving detection accuracy. The model also generates probabilities for each bounding box for each category. These probability values are used to determine the category of the target, so as to achieve multi-category target detection[2].

4. Verify

4.1 Verification and analysis of YOLOv5 model

environment	Distance range (meters)	Measurement error range (%)	Typical distance (meters)	Typical error (meters)
indoor	1-5	Within 5%	3	About 3%
outdoor	0.5-10	Within 8%	5	About 6%
complex	0.2-10	Within 10%	3	About 8%

Figure 1 Experimental results

In the experiment, we used the intelligent guide glasses based on YOLOv5 to conduct binocular ranging experiments on multiple objects. The experimental results show that the intelligent guide glasses can calculate the distance between the object and the camera more accurately. The following is the specific data of the experimental results.

4.1.1 Target detection data set

There are 7787 pictures in the training set, 2317 pictures in the test set, and 1832 pictures in the verification set for target detection, as shown in Figure 2, taking traffic lights as an example.

Class name	Training set		Test set		Validation set	
	Graphics	Label box	Graphics	Label box	Graphics	Label box
red light	650	1220	150	400	100	365
green light	600	1210	150	483	100	358
person	1000	11321	300	516	250	416
bicycle	500	700	150	350	100	297
car	600	850	150	265	100	150
motorbike	700	1600	200	500	150	413
bus	750	5487	256	2566	200	2105
reflective cone	784	1210	261	404	200	315
truck	1550	3530	600	891	552	747
warning column	653	905	100	351	80	214

Figure 2 Data set of target detection

4.1.2 Anchor frame treatment

The K-means clustering algorithm uses distance as a metric to classify and process data sets. The steps are as follows: 1. Randomly set k samples in the sample as the initial clustering Ck; 2. Calculate the shortest distance between the remaining samples of the sample set and the cluster center, and divide them into clusters to which the nearest cluster center belongs; 3. Repeat the steps of 2, and then update k cluster center positions until the cluster center no longer changes its position or reaches the specified number of iterations.

```
anchors:
  - [10,13, 16,30, 33,23]  # P3/8
  - [30,61, 62,45, 59,119]  # P4/16
  - [116,90, 156,198, 373,326]  # P5/32
```

Figure 3 Anchor box cluster size

In view of the data set containing targets of various sizes, in order to adapt to the targets of traffic lights of different sizes, the value of clustering center k is set to 9, and nine anchor boxes are clustered, as shown in Figure 3.

4.2 Deployment verification of YOLOv5 and binocular ranging algorithm

Figure 4 shows the recognition and ranging of obstacles by simulated intelligent guide glasses in a simple environment. The obstacles in the figure are mainly vehicles and people, and the model performs well in the detection of obstacles, with fewer cases of error detection, undetected targets and repeated detection. Moreover, the border accuracy of each target is high, and the actual detection experiment results of the model are good, and the model recognizes two wrong results of close-range targets, and the recognition effect of long-distance targets is good.

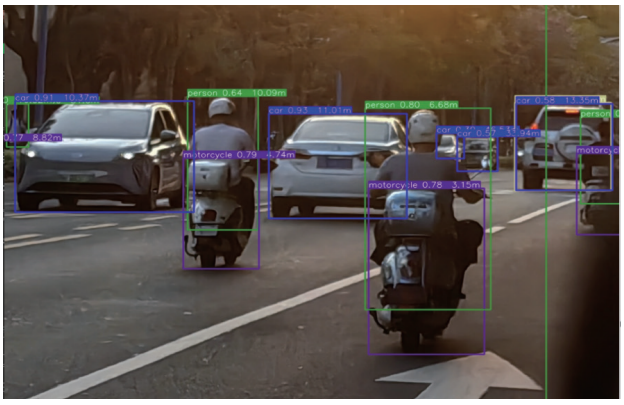

Figure 4 Verifies the detection results on the set

5. Summary

On the whole, in the process of testing the intelligent guide glasses, the intelligent guide glasses can accurately identify obstacles and measure the distance of obstacles, and remind the visually impaired to stay away from obstacles through the offline voice module. When the visually impaired walk at the intersection, through the intelligent guide glasses to recognize the traffic lights to help the visually impaired through the intersection, to ensure the safety of the visually impaired travel.

References:

[1] Elfwing, Stefan, Eiji Uchibe and Kenji Doya. “Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning.” Neural networks: the official journal of the International Neural Network Society 107 (2017): 3-11

[2] Redmon J , Divvala S , Girshick R ,et al.You Only Look Once: Unified, Real-Time Object Detection[C].Computer Vision & Pattern Recognition.IEEE, 2016.DOI:10.1109/CVPR.2016.91.