Blind guidance system based on image recognition and convolutional neural network

Hao Wang*, Kaijie Jiao
School of software, Zhengzhou University, Henan, China
*Corresponding author: wanghao0129@stu.zzu.edu.cn

Abstract. This paper applies straight-line driving technology and pothole detection technology in the field of unmanned driving and edge detection in image recognition to the blind guide system. Researched and produced an intelligent, popular, safe and reliable intelligent blind guide system. This blind guide system uses Python and OpenCV as the development environment under the Linux operating system to realize image recognition and integrate functions on the Jetson nano development board. Through the simulation and analysis of scenes such as the life and travel of the blind, the designed blind guidance system mainly realizes the following functions: image recognition, traffic light recognition, zebra crossing detection, road pothole detection, voice interaction, and assisting straight-line walking.

1. Introduction
As the world's population increases and the aging process accelerates, global blindness is getting worse. China has the largest population of visually impaired people in the world, with more than 17 million in 2018[1]. Although today's guide equipment has improved the inconvenience for the blind to travel, there are still more or less problems. High-tech products such as smart glasses have high cost, complex use, small audience, users still need to have weak vision, low safety, can not be put into mass production, etc., while ordinary intelligent guide stick has poor interaction, poor real-time performance and other problems [2].

Based on this, this paper creatively applies straight-line driving technology and pothole detection technology in the field of unmanned driving and edge detection in image recognition to the blind guide system. Researched and produced an intelligent, popular, safe and reliable intelligent blind guide system. Through the simulation analysis of scenes such as the life and travel of the blind, the designed blind guidance system mainly realizes the following functions: image recognition, traffic light recognition, zebra crossing detection, road pothole detection, voice interaction.

2. Related theories and technologies

2.1. Image recognition technology

2.1.1. Realization of image recognition technology. The video stream is captured by binocular camera simulating human eyes, and the object detection algorithm yolov3 is used to detect obstacles. In the real-time target detection algorithm, the positioning and accuracy of yolov3 have been greatly improved. At
the same time, yoov3 has better universality, which is very suitable for our intelligent blind guidance system.

Basic principle: first extract features from the input image through the feature extraction network to obtain a feature map of a certain size, such as 13*13, and then divide the input image into 13*13 grid cells [3], and then if the center of the touch target is located Which grid cell is in, then the grid cell predicts the target. Each grid cell predicts 3 bounding boxes of a fixed number. The predicted output feature map has two dimensions that are the dimensions of the extracted features, such as 13*13, and another dimension is B*(5+C), where B represents the number of bounding boxes predicted by each grid cell (At this time, B=3), C represents the number of categories of the bounding box, and 5 represents 4 coordinate information and a target score.

Category prediction: using multi-label classification, replacing the softmax function [4] with multiple independent logical classifiers to calculate the probability that the input belongs to a specific label. When calculating the classification loss, yolov3 uses binary cross-entropy loss for each label [4]. This also avoids the use of the softmax function and reduces the computational complexity. Bounding box prediction: Yolov3 uses logistic regression to predict the target score of each bounding box. The method of calculating the cost function is changed, and the relative coordinates of the center point of the b-box relative to the upper left corner of the grid unit are predicted. Multi-scale prediction: Yolov3 uses multi-scale to detect objects of different sizes. The finer the grid cell, the finer the object can be detected.

Feature extraction: Darknet-53 replaces Darknet-19 as the feature extractor.

2.1.2. Traffic light recognition. Traffic light recognition is divided into two steps: one is color recognition, the other is shape recognition. The color must be within the red or green confidence interval, and the shape must be circular within the threshold.

Methods: N frames of red light and green light chroma direction images were taken to construct the red light and green light model, and the stable confidence interval was obtained by calculation, so as to realize the traffic light state recognition. 90% of the data of regional normal distribution is taken as the chromaticity characteristics of red light and green light respectively to identify red and green. The confidence interval of red light is as follows

$$[x_1, x_2] = [\mu_R - 1.65\sigma_R, \mu_R + 1.65\sigma_R]$$

(1)

Where, x is the marginal value of confidence interval; R-red; \(\mu\) - mean; \(\sigma\) - standard deviation.

The green light confidence interval is as follows

$$[x_3, x_4] = [\mu_G - 1.65\sigma_G, \mu_G + 1.65\sigma_G]$$

(2)

Among them, g-green.

Red light when the following conditions are met:

$$\begin{align*}
R_L &= 1 \\
I_o &\leq x_1 & x_2
\end{align*}$$

(3)

Where, \(I_o\) - confidence interval;

\(x_1, x_2\) - red light confidence interval boundary value;

\(R_L\) - judgment result, 1 is true.

Green when the following conditions are met:
Where, \( I_0 \) - confidence interval; 
\( x_3, x_4 \) - boundary value of green light confidence interval; 
\( G_L \) - judgment result, 1 is true.

The circular degree is segmented and judged by the segmentation results:

\[
\begin{aligned}
G_L &= 1 \\
x_3 &\leq I_0 \leq x_4
\end{aligned}
\] (4)

Among them, \( S_c \) - area of regional block; \( p \) - perimeter of block.

If the following conditions are true, the result is an effective circle

\[
\begin{aligned}
Z_c &= 1 \\
S_c &\geq T_c
\end{aligned}
\] (5)

Where, \( Z_c \) - judgment result, 1 is true; \( T_c \) - circular threshold.

2.1.3. Zebra crossing recognition. Pedestrian crossing the road in accordance with the zebra crossing planning route, we should be able to ensure that the blind can at the right time, in the zebra crossing within the scope of the provisions of the safety of the road. The former can judge the right time through traffic light recognition technology, while the latter involves the recognition of zebra crossing, which can not completely copy the technology of traffic light recognition.

In the process of walking, due to the angle and height of the camera, the pictures often can not contain a complete zebra crossing. The exception of the zebra crossing has white pollution inside, which can not accurately identify the zebra crossing as the traffic lights.

We need a set of algorithms based on line extraction and image enhancement to assist the recognition of zebra crossing. After processing the zebra crossing information detected by the camera of the blind guide, we can know that the blind are safe to pass the road through the way of voice prompt.

2.1.4. Zebra crossing feature data extraction. In order to extract the zebra crossing feature data, the edge of the zebra crossing should be extracted by edge detection. After edge extraction, the zebra crossing feature is obvious, which is convenient for Hough transform [5].

Here we use the Canny operator for edge detection, which is accurate in positioning and less affected by noise, which can facilitate the identification of zebra crossings. Figure 1 shows the result of edge detection:

![Figure 1. Edge detection result](image)
After edge extraction, the feature of zebra crossing is a group of parallel lines with decreasing distance and length. Hough transform transforms the information of parallel lines into Hough space for detection, and obtains the feature data of zebra crossing.

The definition of Hough transform in two-dimensional space is as follows

\[
R(\rho, \theta) = \iint f(x, y)\delta(\rho - x\cos \theta - y\sin \theta)dxdy
\]  

(7)

In the formula: \(F(x, y)\) is the gray level of the image point \((x, y)\); \(\rho\) is the vertical distance from the coordinate origin to the line; \(\theta\) is the angle between the vertical line of the line and the X axis.

### 2.1.5. Zebra crossing identification

Through the zebra crossing extraction operation, the edge line feature data of the image is obtained. After Canny edge detection, the zebra crossing feature is obvious compared with other things, which is a group of parallel lines. It can be seen that the distance between parallel lines from near to far is from large to small, that is, the width spacing of these parallel lines shows a decreasing trend. A mathematical model is established by simulating the height change of the camera and calculating the reduction between the remote horizontal lines. If the change of the distance between the detected horizontal lines in the process of walking conforms to this mathematical model, the detected horizontal line can be considered as a zebra crossing.

### 2.2. Road pothole detection technology

Use OpenCV for pothole detection. Divided into 4 steps, first blur the image, then set a threshold for the image to highlight potholes and clean roads; Canny algorithm draws the contour of the edge on the processed image. Finally, the algorithm faster_rcnn is used to detect the contour to determine whether it is a pothole. Let's introduce it in detail next.

Manipulate each pixel of the image. There are two ways to make random pixels have another value, one is to directly change the points in the matrix, and the other is to use the kernel. The kernel is a small matrix with a certain value, which can be used as a filter superimposed on the image. Through the convolution of the image and the kernel, the value of random pixels can be changed to achieve the blur of the image.

Given an image, draw its histogram and choose a value as the threshold. Every pixel larger than this value will become black, and every pixel smaller than this value will become white.

The edge detection algorithm finds the edge in the image. The Canny edge detection algorithm is used here, which will detect the edges of the image and output an image with only contours.

The convolutional neural network model inception_v1 is used as the basis for the classification of potholes in multiple scenes, and the algorithm faster_rcnn is used for corresponding detection [6]. The training process of deep neural networks is carried out in batches. To maximize the training results, this algorithm Use the momentum optimization method Momentum to optimize it:

\[
v_w = \beta v_w + (1 - \beta) w,
\]  

(8)

\[
v_b = \beta v_b + (1 - \beta) b,
\]  

(9)

\[w = w - \alpha v_w,
\]  

(10)

\[b = b - \alpha v_b,
\]  

(11)

Where, \(v_w\) and \(v_b\) represent the weight and bias gradient momentum accumulated in the previous iteration, respectively; \(\beta\) is the gradient accumulation index; W and B are respectively the weight gradient and bias gradient obtained by the reverse propagation of the loss function. Alpha is the learning rate.
3. Conclusions
Through the study and analysis of the guide system, the author designs a voice interactive intelligent guide system that can independently identify road obstacles, traffic lights and zebra crossings, which almost solves the problem of difficulty for the blind to go out and ensure the safety of the blind. It will bring good news to the visually impaired groups in China and make great contributions.

References
[1] Global data on visual impairments 2010[M]. 2012.
[2] Qiu Jingjing. Research on some problems of multifunctional blind guide device[D].Zhengzhou University, 2010.
[3] Yi S,Nie Y,Zhang Y Y.et al.A Method for Nighttime Target Recognition Based on Infrared Thermal Imaging And YOLO v3 Method.vol.41,no.10, pp.972,2019.
[4] Wei H B,ZHANG D J,Du G M.Vegetables Recognition Algorithm Based on Modified YOLO v3.vol.41,no.02,pp.7-8,2020.
[5] Hu J.Research and Implementation of Blind Navigation GLASSES Based on Ultrasonic and Image Recognition.UESTC.2013.
[6] Chen P,Ying J.Multi-scene Road Pothole Image Detection Based on Convolutional Neural Network.vol.49,no.01, pp.97.