参赛学生姓名：张景涵　田桢干

中学：南京外国语学校

省份：江 苏

国家/地区：中 国

指导老师姓名：张 川

指导老师单位：东南大学

论 文 题 目：A Novel Crosswalk Traffic Light Detection Algorithm Resolving Multi-Light Interference for the Visually Impaired

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

A Novel Crosswalk Traffic Light Detection Algorithm Resolving Multi-Light Interference for the Visually Impaired

Jinghan Zhang and Zhengan Tian

Nanjing Foreign Language School, Nanjing, China

Abstract: This paper proposes a novel algorithm for detecting crosswalk traffic lights, specifically designed to aid visually impaired pedestrians. First, we identify five key image characteristics and accordingly summarized the five factors that the algorithm needs to have. Then, using this priori knowledge, we design our algorithm. This algorithm associates crosswalk traffic light with crosswalk, strengthens small target detection, and uses an independent color detection module, which effectively addresses multiple issues, including multi-light interference. In implementation, the algorithm incorporates squeeze-and-excitation (SE) and convolutional block attention module (CBAM) mechanisms to improve the detection of small objects and crosswalks under various conditions. A two-layer CNN is also designed for color detection of virous crosswalk traffic lights. Experimental results show that our algorithm achieves a detection accuracy of 97.5% with score weight (0.5,0.5) and selected area (50, 20), substantially outperforming the 92.0% accuracy of the YOLOv5 combined with the maximum pixel detection method. Sensitivity analysis also reveals that our algorithm has strong robustness to noise. The results show that our algorithm effectively improves the safety of visually impaired individuals while crossing the street.

Keywords: Crosswalk traffic light, Crosswalk, Object detection, Convolutional neural network, YOLOv5, Visually impaired assistance

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

Contents

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

1. Introduction

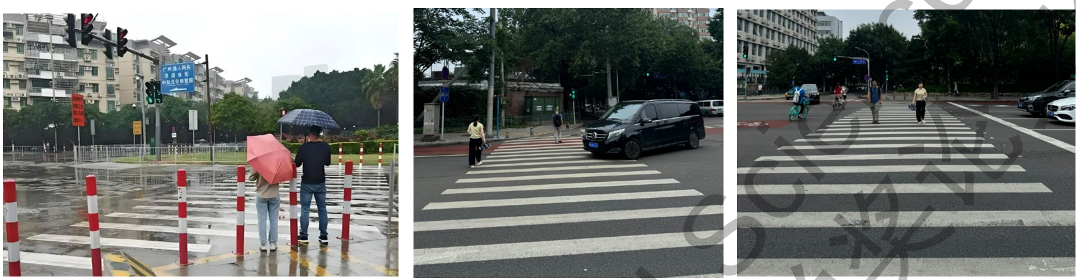
The visually impaired community is easily overlooked. In their daily lives, they often encounter many inconveniences. For example, blind pathways are often obstructed, and most crosswalk traffic lights lack indicators for the visually impaired. Through our communications with visually impaired friends, we discovered that, contrary to our previous assumptions, most of them can use smartphones quite well with the help of voice assistance. They can utilize functions like taking photos and even sharing pictures on social media. Therefore, we decided to leverage the camera function of smartphones to develop an algorithm for detecting crosswalk traffic lights to help visually impaired individuals safely cross the street.

In recent years, since the demand for autonomous vehicles and smart city applications has grown significantly, traffic light detection and recognition technologies have advanced quickly. Traditional traffic light detection approaches mainly rely on classic computer vision techniques, such as color segmentation, edge detection, and shape recognition. However, different lighting conditions, obstructions, and weather conditions often make these approaches less reliable in practical scenarios. In recent years, methods combined with deep learning, especially convolutional neural networks (CNNs)[1], have been proposed to improve the accuracy and robustness of traffic light detection. CNN-based object detection algorithms are generally categorized into two types: two-stage detection algorithms and one-stage detection algorithms[2]. Two-stage algorithms, such as R-CNN[3] series, first input the image into a convolutional neural network (CNN) for feature extraction and generate a sparse set of candidate regions. Then, a second CNN is used to perform targeted feature extraction on these candidate regions. Finally, the system perform classification. One-stage detection algorithms, such as YOLO[4] series and RetinaNet[5], do not require generating pre-selected regions that might contain objects. Instead, they divide the image into a grid (e.g., 10×10 cells) and detect objects directly within each grid cell, significantly reducing the processing time. Compared with two-stage algorithm, one-stage algorithm has faster detection speed but slightly lower accuracy. Recent advances in YOLO-based traffic light detection demonstrate its suitability for real-time applications, especially in autonomous driving. In reference [6], an improved YOLOv4[7] combined with shallow feature enhancement mechanism and the bounding box uncertainty prediction mechanism is proposed to detect traffic light. In reference [8], an YOLOv5[9] algorithm with K-means clustering is used to speed up traffic light detection. Reference [10] proposes a real-time railroad signal light detection method based on YOLOv5 and verifies the effectiveness of the method. In reference [11], an YOLOv8[12] system equipped with GPU accelerators is used to identify real-time traffic lights. In addition, there are many related research works. These researches show that the YOLO series algorithms have good real-time performance in traffic signal identification and also have high accuracy in complex scenarios, with robust traffic signal detection capabilities under various lighting and weather conditions. Therefore, we will propose our crosswalk traffic light detection algorithm based on the YOLO algorithm.

It should be noted that, although current algorithms have achieved good detection performance in the field of traffic light detection, they still cannot solve crosswalk traffic light detection problem. Most current traffic light detections focus on vehicle traffic lights, paying very little attention to crosswalk traffic lights[13-15]. Unlike vehicle traffic light detection, crosswalk traffic light detection not only needs to identify smaller targets, but also needs to detect the right crosswalk traffic light from multiple traffic lights in one image. In figure 1, there are multiple traffic lights in each image. For example, in figure 1(a), the crosswalk traffic light is located to the left of the blue umbrella and

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

is quite small and difficult to identify. The other two vehicle traffic lights in this image are more obvious, and because the colors are different, it is very easy to misjudge. At present, there is no research work that simultaneously considers these two challenges: selecting the right light from multiple lights and small targets detection. In the analysis of subsequent sections, we also observe some other characteristics of crosswalk traffic lights, such as the diversity of types, which have not been considered in existing studies. Such as in references [13-15], although the crosswalk traffic lights are identified, they are not distinguished from the traffic lights of vehicles or other crosswalk traffic lights in the same image. As a result, these methods may lead to detection errors and cause safety problems.



(a) (b) (c)

Figure1. Images of crosswalk traffic light

In this paper, we propose a novel crosswalk traffic light detection algorithm with anti-multi-light interference capability for the visually impaired. When there are multiple traffic lights in an image, the crosswalk traffic light will be detected based on its position relative to the crosswalk and the area of the traffic light to resolve multi-light interference. Two attention mechanisms are introduced: Squeeze-and-excitation (SE)[16], which is used for detecting small targets, and convolutional block attention module (CBAM)[17], which is used for detecting crosswalk. At the same time, a two-layer CNN is specifically designed for color detection. The main contributions of this paper are as follows:

1. Our work aims to help visually impaired people cross streets safely by accurately detecting crosswalk traffic lights. We conduct a detailed analysis of crosswalk traffic light images, extracting key characteristics and identifying essential factors required for an effective detection algorithm. Building on priori knowledge, we design a solution specifically for crosswalk traffic light detection. To our knowledge, this is the most comprehensive approach currently available for crosswalk traffic light detection.
2. To solve the problem of misidentification caused by multiple traffic lights, the designed algorithm associates crosswalk traffic light with the corresponding crosswalk location. It effectively improves the detection accuracy and makes crossing streets safer for the visually impaired.
3. To enhance the detection accuracy for small targets, the SE model is introduced after C3 block at the neck of YOLOv5. By adaptively adjusting the importance of each feature channel, the SE module greatly enhances traffic light detection ability and enables it to detect crosswalk traffic light even if it only occupies a small area in the image.
4. To detect the crosswalk, CBAM is added before each C3 block at the neck of YOLOv5 and after SPPF block at the backbone of YOLOv5. By introducing channel attention and spatial attention, the model can filter and enhance features before convolution, resulting in better detection accuracy.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

5. To enhance color recognition for different types of crosswalk traffic lights and improve safety, we design a two-layer CNN for color detection. Experimental results show that this module effectively identifies the colors of various traffic light types and has strong robustness.

The remainder of this article is organized as follows. Section 2 provides the preliminaries of this work. The proposed crosswalk traffic light detection algorithm is detailed in Section 3. Section 4 presents the experimental results and performance analysis. Finally, section 5 concludes the article and gives future research directions.

2. Preliminaries

This section introduces the essential preliminaries related to our algorithm, including CNN and evaluation metrics.

2.1 CNN

In the field of image detection, a convolutional neural network is typically composed of convolutional layers, pooling layers, and fully connected layers. It performs classification by using a structure with multiple connected layers. A typical structure of CNN is shown in figure 2.

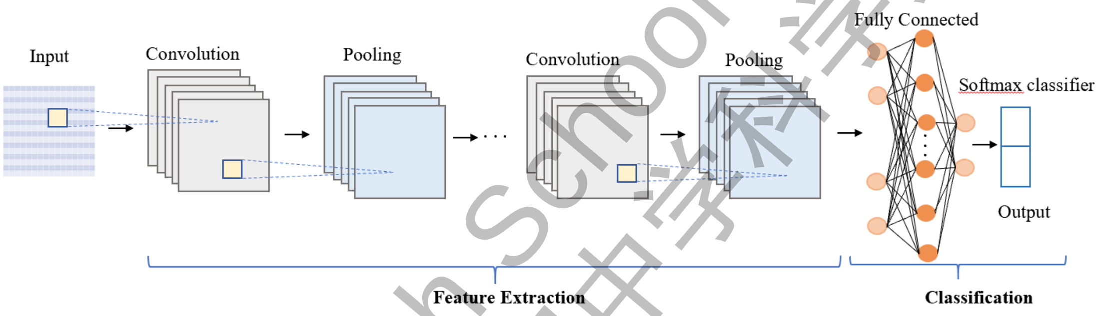


Figure 2. The structure of CNN

The primary function of the convolutional layer is to extract features presented in the previous layer. It uses convolutional kernels of different sizes (like 2×2 or 3×3) to capture local patterns. The convolutional kernel operates on the input image or feature map, and the results are passed through an activation function to enhance nonlinearity. The activation function is placed at the end or between layers of neural networks and transforms the input to keep its values within a manageable range. Common activation functions include the Sigmoid function, ReLU function, etc.. This process can be expressed as

$$C^l = f(A^l * W^l + B^l) \tag{1}$$

where C^l is the output of l layer, A^l is the input of l layer, W^l is the learnable weight vector, B^l is the bias vector, $f(\,)$ is the activation function, and * is the convolutional operation.

In CNN, a convolutional layer is followed by a pooling layer. Pooling layer, also known as downsampling layer, is used to reduce the size of the input feature map and increase the receptive field of the following convolutional layers, reducing overfitting in the network. A pooling operator integrates the data from a small area (like a rectangle) into a single value. The most popular pooling methods are max pooling and average pooling. Figure 3 presents the process of convolution and pooling.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

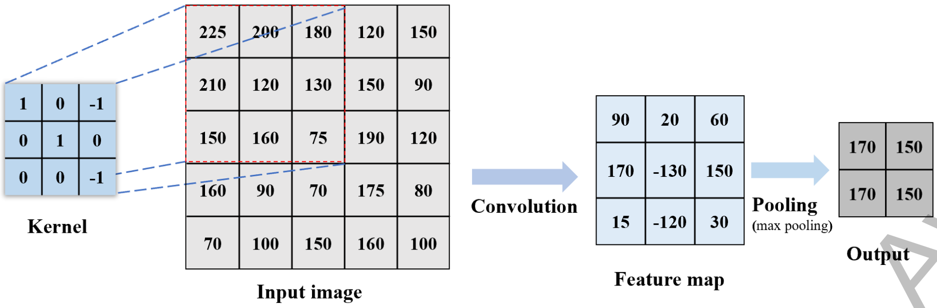


Figure 3. Example of the process of convolution and pooling (max pooling)

The fully-connected layer synthesizes the features into the final classification. In the fully-connected layer, each neuron is connected to all the input neurons from the previous layer, with connections governed by weights and biases. It can be expressed as:

$$y_j = f\left(\sum_i (w_{i,j}.x_i) + b_j\right) \quad (2)$$

where $w_{i,j}$ is the weight between neuron i and neuron j, x_i is the input from neuron i, y_j is the output of neuron j, b_j is the bias for neuron j, and $f(\ )$ is also the activation function.

At the end, the Softmax function is applied in the output layer to normalize raw scores into a probability distribution, ensuring that the sum of probabilities across all classes equals 1.

2.2 Evaluation metrics

Confusion matrix is a table used to evaluate the performance of a classification model by comparing its predicted labels with the actual labels, as shown in figure 4.

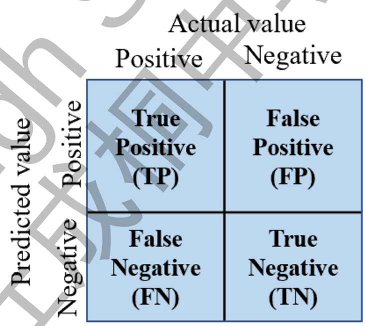


Figure 4. Confusion matrix

Here, true positive (TP) is the number of correctly predicted positive classes by the model, true negative (TN) is the number of correctly predicted negative classes by the model, false positive (FP) is the number of incorrectly predicted positive classes by the model, and false negative (FN) is the number of incorrectly predicted negative classes by the model. From confusion matrix, some evaluation metrics are defined.

Accuracy is the proportion of correct predictions made by the model out of all predictions:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision is the proportion of the predicted positive results that are actually positive:

$$\text{Precision}=\frac{TP}{TP+FP} \quad (4)$$

Recall is the proportion of actual positives that are correctly identified by the model:

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

$$\text{Recall}=\frac{TP}{TP+FN} \tag{5}$$

mAP is the average of the average precision (AP) across all classes, where AP is the area under the precision-recall curve for a given class:

$$\text{AP}=\int_0^1 p(r)dr \tag{6}$$

where $p(r)$ is the precision as a function of recall.

$$\text{mAP}=\frac{1}{N}\sum_{i=1}^{N}\text{AP}_i \tag{7}$$

where AP_i is the average precision for class i.

IoU (Intersection over Union) is a commonly used metric to measure the overlap between the predicted bounding box and the ground truth bounding box. It is calculated as follows:

$$\text{IoU}=\frac{\text{Intersection Area}}{\text{Union Area}} \tag{8}$$

where intersection area is the overlapping area between the predicted box and the ground truth box, and union area is the total area covered by both the predicted and ground truth boxes, subtracting the intersection area.

3. Implementation

3.1 System framework

To achieve better detection results, we take many photos from the perspective of pedestrians crossing the street. By analyzing the images, we summarize **five characteristics** of this type of images:

(1) Multiple traffic lights often appear in one image;

(2) In some images, the crosswalk traffic light is smaller than other vehicle traffic lights;

(3) Crosswalk traffic light always appears together with the crosswalk;

(4) The diversity of the relative positions between crosswalk traffic light and crosswalk;

(5) There are many types of crosswalk traffic lights, with varying display style.

According to these characteristics, the detection algorithm needs to incorporate the following **five factors**:

(1) The algorithm needs to have the ability to eliminate interference from multiple lights and detect the crosswalk traffic light;

(2) The algorithm should be able to accurately detect small objects, especially when crosswalk traffic lights are smaller than vehicle traffic lights;

(3) The algorithm needs to associate traffic lights with related objects like crosswalks and distinguish between pedestrian and vehicle traffic lights;

(4) The algorithm must recognize and adapt to the diverse relative positions between crosswalk traffic light and crosswalk;

(5) The algorithm needs to accurately detect the color of different types of crosswalk traffic lights.

Existing image detection algorithms, such as the YOLO series and RetinaNet, cannot fully meet the above requirements. These algorithms treat each object independently and may fail to understand that the presence of a crosswalk increases the likelihood that the nearby traffic light is

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

the crosswalk traffic light. They might confuse crosswalk traffic lights with vehicle traffic lights, especially when multiple lights are present.

Therefore, we propose a new algorithm for crosswalk traffic light detection. Considering that we hope to run the algorithm on mobile devices, YOLOv5 and YOLOv8 are both viable options. After analysis and comparison, YOLOv5 is chosen for its advantages. YOLOv5 has proven stability across different platforms, which is crucial for ensuring safety in real-world applications. It also offers an excellent balance between speed and accuracy, making it well-suited for real-time tasks. Its rich APIs and multiple versions allow for easy deployment, particularly in resource-constrained environments like mobile devices. Additionally, its active community provides extensive third-party tutorials, tools, and support, which are valuable for development. However, considering that its accuracy might be slightly inferior to YOLOv8, two attention mechanisms are added to ensure detection accuracy.

Based on the characteristics of these images and the required factors of the detection algorithm for crosswalk traffic light, a module functional diagram of the designed novel algorithm is shown in Figure 5. After inputting the image, the improved YOLOv5 algorithm with the SE attention module is first used to detect the crosswalk traffic light. If only one traffic light is detected, the detected traffic light will be sent to color detection module. If multiple traffic lights are detected in the image, the image is reprocessed into the improved YOLOv5 algorithm with the CBAM attention module to detect the crosswalk. Based on the detected crosswalk, a nearby area will be selected, and the SE combined with YOLOv5 algorithm will be used again to detect crosswalk traffic in the selected area. Finally, the detection results will be sent to color detection module to detect the colors of different types of crosswalk traffic lights.

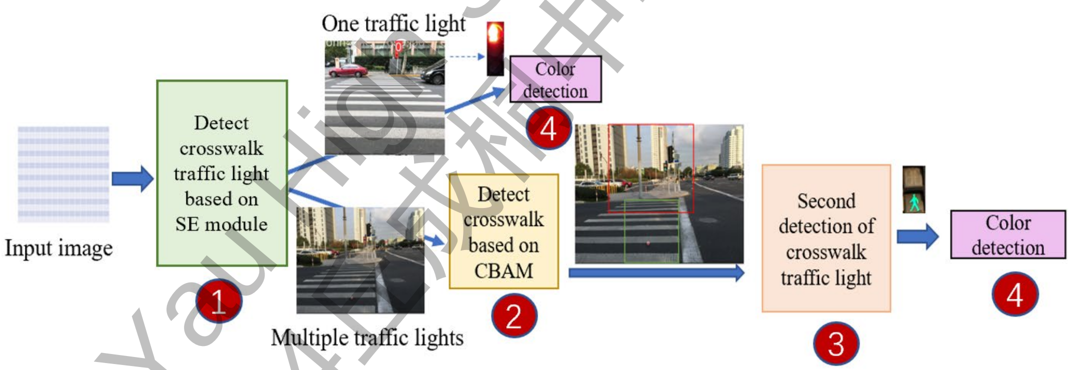


Figure. 5 Module functional diagram of the designed crosswalk traffic light detection algorithm

The functions of the **four modules** in the system framework are as follows:

(1) Detect crosswalk traffic light based on SE module: The SE module is used in YOLOv5 to detect the traffic lights and improve the detection of small objects, specifically the smaller crosswalk traffic light.

(2) Detect crosswalk based on CBAM: After detecting multiple traffic lights, the crosswalk needs to be detected to assist in locating the crosswalk traffic light. This module incorporates the CBAM attention mechanism into the YOLOv5 to detect crosswalk and enhance detection accuracy.

(3) Second detection of crosswalk traffic light: After detecting the crosswalk, this module performs a second traffic light detection in the selected area to exclude interference from multiple lights and identify the crosswalk traffic light.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

(4) Color detection: To improve color detection accuracy for different types of crosswalk traffic lights and enhance system safety, the algorithm incorporates a dedicated module specifically for color detection. This module uses a specialized CNN model to classify the color (red or green) of the traffic light in the cropped image area.

The above modules can be used to address the five essential factors required for this detection algorithm, with the corresponding relationships shown in table 1. The detailed implementation of each module is described below.

Table 1. Corresponding relationships between modules and features required for the algorithm

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Module 1		✓			
Module 2	✓		✓	✓	
Module 3	✓		✓	✓	
Module 4					✓

3.2 Detect crosswalk traffic light based on SE model

The traffic light detection requires the model to accurately identify small and specific objects. The SE module can adaptively assign weights to each channel, enhancing the attention on important feature channels. The SE module is added after the C3 module at the neck of YOLOv5. By the squeeze operation, it compresses global information into a single value for each channel, and then reallocates the importance of each channel through a fully connected layer by excitation operation. This process can dynamically adjust the responses of individual channels based on global information, enhancing the characteristic information of small targets.

By the squeeze operation, global spatial information from each channel is compressed into a single value using global average pooling. This operation effectively squeezes the spatial dimensions of the feature map. This operation can be formulated as

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \tag{9}$$

where $x_c(i,j)$ is the value at spatial position (i,j) of the *c-th* feature map, H and W are the height and width of the feature map, and z_c is the squeezed scalar for channel c.

Excitation operation is after squeezing operation and produces channel-wise weights that scale the importance of each channel. This operation passes the global pooled information through two fully connected (FC) layer with a non-linear activation function ReLU followed by a Sigmoid activation. This operation can be formulated as:

$$s_c = \sigma(W_2 \delta(W_1 z)) \tag{10}$$

where W_1 and $W2$ are the weights of the two FC layers, δ is the ReLU activation function, σ is the Sigmoid activation function, and s_c is the final weight for the *c-th* channel.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

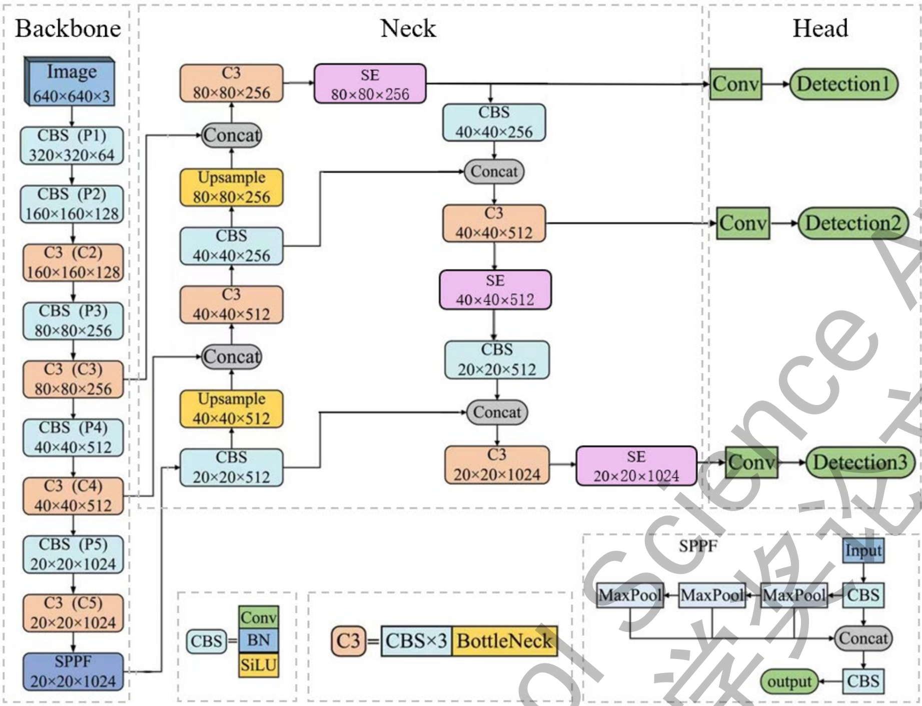


Figure 6. Architecture of improved YOLOv5 with SE model for traffic light detection

The remaining blocks in Figure 6 are:

(1) Backbone

The backbone of the network is responsible for extracting features from the input image. It consists of the following components:

• Image Input: A 640×640 pixel image with 3 color channels (RGB).

• CBS Blocks: Convolutional layers followed by Batch Normalization and Sigmoid Linear Unit (SiLU) activation.

• C3 Blocks: Blocks consisting of three CBS blocks followed by a residual connection (skip connection).

• SPPF Blocks: Module that aggregates contextual information by using multiple max-pooling operations of different kernel sizes, which has an output size of 20×20×1024.

(2) Neck

The neck is responsible for fusing features from different scales (feature pyramid) and enhancing the model's ability to detect objects of different sizes.

• Upsample and Concat: Features from different layers are upsampled and concatenated with higher-resolution feature maps:

– The 20×20×1024 features from the SPPF block are concatenated with the 20×20×512 features from the CBS block, followed by a C3 block to output 20×20×1024 features.

– Output of the CBS block is further upsampled to 40×40×512, concatenated with the 40×40×512 features from C3, and processed through another C3 block to output 40×40×512 features.

– The output is again upsampled to 80×80×256, concatenated with 80×80×256 features from C3, and passed through a C3 block to produce 80×80×256 features.

(3) Head

The head consists of three detection layers that predict bounding boxes, objectness scores, and

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

class probabilities at three different scales:

- Detection1: Operates on 80×80×256 features for detecting small objects.
- Detection2: Operates on 40×40×512 features for detecting medium-sized objects.
- Detection3: Operates on 20×20×1024 features for detecting large objects. Each detection layer consists of a convolutional layer that outputs predictions.

By adopting this improved YOLOv5 with SE model, the crosswalk light will be detected, but other traffic lights will also be selected. Therefore, the image will be sent to the "Detect crosswalk based on CBAM" module, as shown in figure 5. In contrast, if there is only one traffic light in an image, the bounding box of traffic light will be cropped, and the cropped image will be sent to the color detection module directly.

3.3 Detect crosswalk based on CBAM

For images with multiple traffic lights, the crosswalk will be used to help locate the crosswalk traffic light to resolve multi-light interference. An improved YOLOv5 algorithm based on CBAM is designed to detect crosswalk.

CBAM is an attention mechanism that can be added to convolutional neural networks to enhance feature representation by focusing on important channels and spatial locations. We adopt CBMA to YOLOv5 for crosswalk detection to improve the detection accuracy for crosswalks by focusing on the most important features in the image. CBAM does this through two types of attention: channel attention (CA) and spatial attention (SA).

Channel attention uses global average pooling and global max pooling to figure out which channels are important and helps the model focus on the most important feature channels. With channel attention, the model can give more weight to the channels that are most relevant to detecting crosswalks, ignoring unnecessary background noise. The channel attention can be formulated as

$$M_c(F) = \sigma(W_1(W_0(AvgPool(F))) + W_1(W_0(MaxPool(F)))) \tag{11}$$

where F is the input feature map, AvgPool and MaxPool extract global features from each channel, W_0 and W_1 are weights in the network, σ is the Sigmoid function.

Spatial attention makes the model concentrate on the most important areas where crosswalk lines are located, especially when those lines might be partially hidden or difficult to see. The spatial attention can be formulated as

$$M_s(F) = \sigma(f^{7\times7}([AcgPool(F); MaxPool(F)])) \tag{12}$$

where $f^{7\times7}$ represents a convolution operation with the filter size of 7× 7, σ is the Sigmoid function, and $[AcgPool(F); MaxPool(F)]$ combines the results of average and max pooling.

Therefore, as shown in figure 7, we add CBAM after SPPF at the tail of the backbone, which enhances the network's ability to select important features across channels and spatial regions. The SPPF module has already helped to capture features at different scales, but adding CBAM after SPPF ensures that the most relevant scale-specific features are emphasized, so that the multi-scale information fused by the SPPF module can be weighted and filtered before entering the Neck. This means the model is better at detecting crosswalks regardless of their sizes or orientations in the frame. We also insert an CBAM module before each C3 block in the neck, enabling the C3 modules to further perform convolution and fusion on enhanced key features. This is particularly helpful in the neck of the YOLOv5 architecture, where features are aggregated from different layers (backbone

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

and previous neck layers).

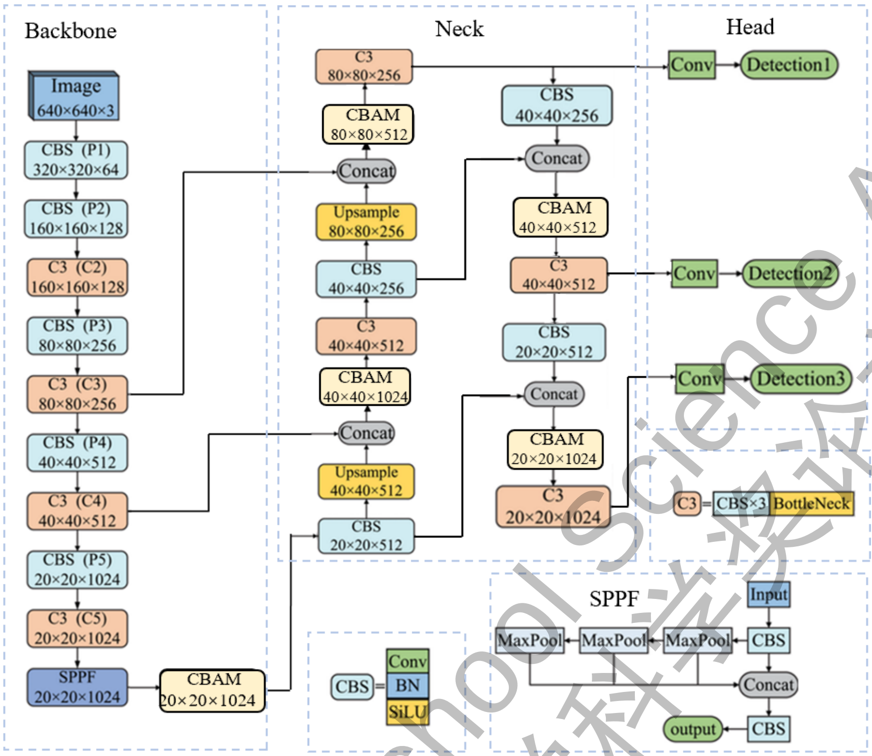


Figure 7. Architecture of improved YOLOv5 with CBAM for crosswalk detection

Figure 8 is the detection results. We can see the crosswalks are marked with green boxes. By labeling the training set and refining the model, the detection results are restricted to crosswalks made up of horizontal lines, corresponding to the current street crossing. This helps avoid interference from crosswalks on other streets.

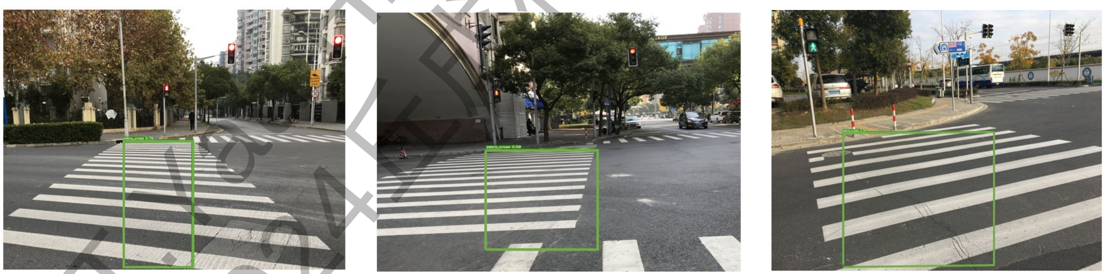


Figure 8. The images marked with crosswalk bounding boxes

3.4 Second detection of crosswalk traffic light

For situations with multiple traffic lights, after detecting the crosswalk, we will select a specific area around the crosswalk and perform the second detection of the crosswalk traffic light in this area. The detection method still uses the method in section 3.2. When selecting the detection area, we need to consider the fourth characteristic of image: The diversity of the relative positions between crosswalk traffic light and crosswalk. Crosswalk traffic lights are generally located at the upper left, upper right, or directly above the crosswalk, as shown in figure 9.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

(a) Upper left (b) Above (c) Upper right

Figure. 9 The diversity of the relative positions between crosswalk traffic light and crosswalk

Meanwhile, we analyze the output images from the crosswalk detection module which are marked with crosswalk bounding boxes. We find that the crosswalk lights are generally located near the crosswalk bounding boxes. Because of different camera angles and the position of the bounding boxes, most traffic lights are positioned directly above or slightly diagonally above the top line of crosswalk bounding. However, we cannot completely rule out situations where the traffic lights may be located below.

Therefore, we use the following cropping function to select the detection area. The cropping function expands the bounding box by m% horizontally, while the vertical cropping is adjusted to include the area from the top of the image to n% below the top boundary of the crosswalk's bounding box. Most of the target traffic lights are exactly above the crosswalk, but some are outside this range. The m% expansion in the horizontal direction is intended to broaden the search area for potential traffic lights around the detected crosswalk, which increases the likelihood of capturing targeted traffic lights that are not on the top of the bounding box. Expanding the bounding box downwards by n% beyond the top boundary of the crosswalk's bounding box allows us to account for variations in traffic light placement height. This small expansion ensures that even if the traffic lights are positioned a bit lower than the bottom of the crosswalk, they will still be captured within the cropped area. Figure 10 shows how to determine the selected area.

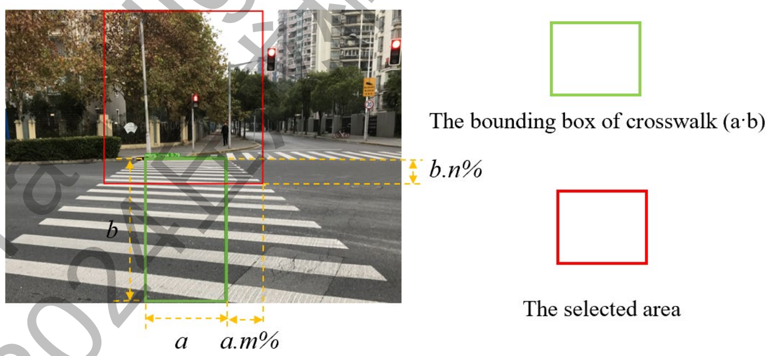
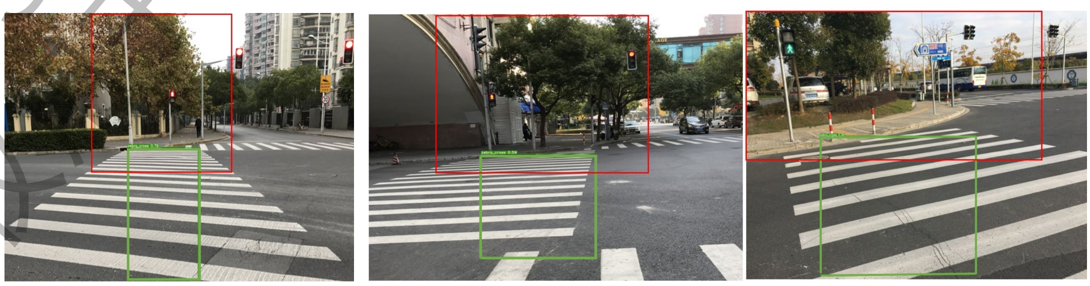


Figure 10. How to determine the selected area

Figure 11 are the examples of selected area.



(a) (b) (c)

Figure 11. The selected area (in red box) where m=50, n=20

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

If multiple traffic lights are still detected within this selected image, such as in figure 10 (b) and (c), we choose the most relevant traffic light using the following method. A score S is calculated for each detected traffic light based on two factors: distance factor D_i is the normalized distance from the i-th traffic light's center to the midpoint of the upper boundary of the crosswalk's bounding box, and area factor A_i is the normalized area of the i-th traffic light's bounding box. The distance factor D_i is calculated as follows:

$$D_i = \frac{{d_i}^2}{{d_{\max}}^2} \qquad i \in \{1, 2, ..., N\} \tag{13}$$

where d_i is the distance from the i-th traffic light's center to the midpoint of the upper boundary of the crosswalk's bounding box, $d_{\max}$ is the maximum value of all d_i, and N is the number of traffic lights detected in the photo.

The area factor A_i is calculated as follows:

$$A_i = \frac{A_i^{'}}{A_{\max}} \qquad i \in \{1, 2, ..., N\} \tag{14}$$

where A' is the area of traffic light's bounding box, and $Amax$ is the maximum area of all A_i'.

Then, the score of the i-th traffic light is defined, and this value is positively correlated with the probability that the traffic light is the target crosswalk traffic light. The score S_i is formulated as

$$S_i = \alpha(1 - D_i) + \beta A_i \tag{15}$$

where α and β are the weights of score for the distance factor and area factor, respectively, and they satisfy $0 \le \alpha \le 1$, $0 \le \beta \le 1$, and $\alpha+\beta=1$.

The traffic light with the largest S_i is selected as the crosswalk traffic light:

$$S_{\max} = \max(S_i) \qquad i \in \{1, 2, ..., N\} \tag{16}$$

By selecting the maximum S_i, we effectively choose the most relevant traffic light for the given crosswalk.

Finally, after the second detection of crosswalk traffic light, the appropriate traffic light is selected. We crop it out from the image and save it for further processing. In our algorithm, the cropped image will be sent to the color detection module.

3.5 Color detection

Based on the fifth characteristic of the images, there are many types of crosswalk traffic light. Figure 12 shows some different types of crosswalk lights in practice. By simulation, we find that some of the crosswalk lights, such as the lights in figure 12 from (4) to (7), are very hard for YOLOv5 to directly detect their color, especially those with extremely high or low brightness and green lights with yellow countdowns.

(1) (2) (3) (4) (5) (6) (7)

Figure 12. Different types of traffic lights.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

Therefore, a two-layer CNN is designed to detect the colors of different types of crosswalk traffic lights, which also satisfies the fifth factor required by the algorithm. The structure of two-layer CNN for color detection is shown in figure 13. First, the input image is resized to 32×32 pixels and normalized, where pixels were scaled to the range [-1,1] from [0,1] in each of the three RGB channels to help the model converge faster when learning. As the pixels are resized to a small scale of 32×32 pixels, both efficiency and feature retention are guaranteed. Second, the resized image is sent to the convolution and activation layers. The first convolution layer uses 32 filters with a kernel size of 3×3, a stride of 1, and padding of 1. The second convolution layer employs 64 filters, also with a 3×3 kernel, a stride of 1, and padding of 1. Both convolution layers are followed by a max pooling layer with a kernel size of 2, a stride of 2, and no padding which selects the maximum value as output when the kernel shifts. By focusing on the most dominant features, the layer effectively reduces unnecessary noises of the cropped traffic light (such as object obstruction in the city and light changes). Third, the multi-dimensional output is sent to the flattening and fully connected layers. A flattening step converts the multi-dimensional output to a one-dimensional array suitable for processing by fully connected layers. The first fully connected layer connects this flat output to 128 neurons, followed by a dropout layer. The dropout layer randomly sets a fraction of the input units to zero with a dropout probability of 0.5. Since traffic lights captured in the city may be affected by many factors, we apply this layer to simulate noises to the model, which enhances model robustness and prevents overfitting. Finally, the dropout layer connects a second dense layer for binary classification, connecting 2 output neurons before the Softmax layer.

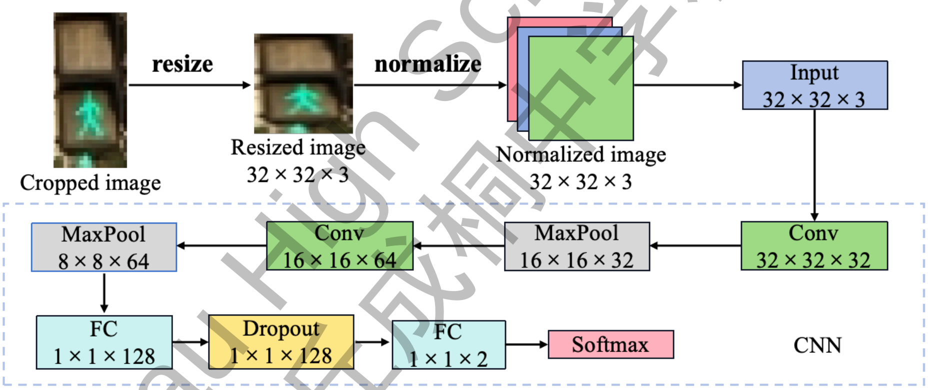


Figure 13. The structure of two-layer CNN for color detection

4. Experiments

4.1 Dataset collection and processing

To train and use our algorithm effectively, the dataset needs to be organized in a specific format. We made the labels on the website makesense.ai[18], an open-source annotation tool that enables drawing bounding boxes around objects in images to label them for our training tasks.

The images used are partly taken by ourselves and partly from ImVisible dataset on GitHub. For our specific goal of crosswalk traffic light and crosswalk detection, we select 600 images from these two sources. These images include crosswalk traffic lights in various scenarios. In the labeling process, we annotate the crosswalk traffic lights and the crosswalks respectively. For the crosswalk traffic lights, we use a single class label. Each traffic light in the image is manually annotated with a bounding box. For crosswalks, the labeling process involves more detailed considerations. As shown in Figure 14, we define the bounding box for crosswalks by selecting the upper edge of the

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

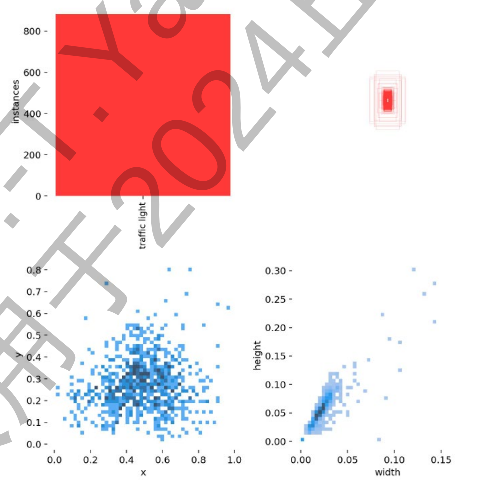
crosswalk as one side of the rectangle and extending it downward to encompass the entire crosswalk area. This approach ensures that the bounding box accurately reflects the shape and orientation of the crosswalk, which can vary depending on the angle and perspective from which the image is taken.

Figure 14. Crosswalk labeling

After labeling, the images and their corresponding annotations are divided into three subsets: training, validation, and test sets. The allocation ratio of training, validation, and test sets is 7:1:2. This distribution ensures that the model is trained on a sufficient variety of data, while also allowing for robust evaluation and tuning during the validation and testing phases.

4.2 Analysis of dataset features

As shown in figure 15(a), the spatial distribution of traffic lights across the dataset is high. Most of the objects are located close to the center, but at the same time, some objects are distributed in a sporadic way, which is subject to the differences in distances, perspectives, and angles. The size distribution of traffic lights emphasizes that most of the objects are small in size, with a very small width and height relative to the size of the image.



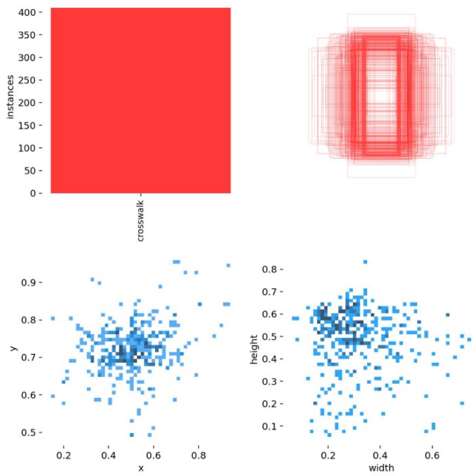


Figure 15. (a) Traffic light dataset feature (b) Crosswalk dataset feature

In figure 15(b), the y values of the bounding box center coordinates, concentrated between 0.7

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

and 0.8, indicate that most crosswalks are located in the lower portion of the images, leaving sufficient space above for selecting the right traffic lights we want. Additionally, the height values, primarily between 0.5 and 0.6, show that crosswalks are tall enough to effectively exclude traffic lights that are invalid in our detection.

4.3 Performance of traffic lights and crosswalk detection

Table 2 shows that the accuracy of the traffic lights detection model using SE module and the crosswalk detection model using CBAM is significantly higher than that of the original YOLOv5s model. In addition, mAP@0.5 has a high value. However, the recall rate is slightly lower. This suggests that these models detect fewer overall targets, but the detected targets are identified with very high accuracy. In contrast, the unmodified original model detects a higher number of targets but with lower accuracy. Since the goal is to help visually impaired people cross streets, it is vital to avoid false instructions. While YOLOv5 without SE has a higher recall rate, it can generate more false positives, which can be dangerous for visually impaired users. This highlights the advantages of the improved model.

Table 2. Performance metrics for traffic light and crosswalk detection

	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Traffic Light YOLOv5s	0.911	0.954	0.967	0.611
Traffic Light YOLOv5s+SE	0.951	0.932	0.968	0.587
Crosswalk YOLOv5s	0.975	0.976	0.971	0.813
Crosswalk YOLOv5s+CBAM	1.000	0.975	0.975	0.779

As shown in figure 16, the loss of the model decreases very fast. Especially in the initial training stage, both box_loss and obj_loss decrease rapidly, indicating that the model is rapidly converging. mAP@0.5 is significantly higher than mAP@0.5:0.95, indicating that the detection performance of the model is better when the threshold of IoU is set to 0.5. Therefore, in this case, setting the IoU threshold to 0.5 is a more appropriate choice, because it can bring higher detection accuracy.

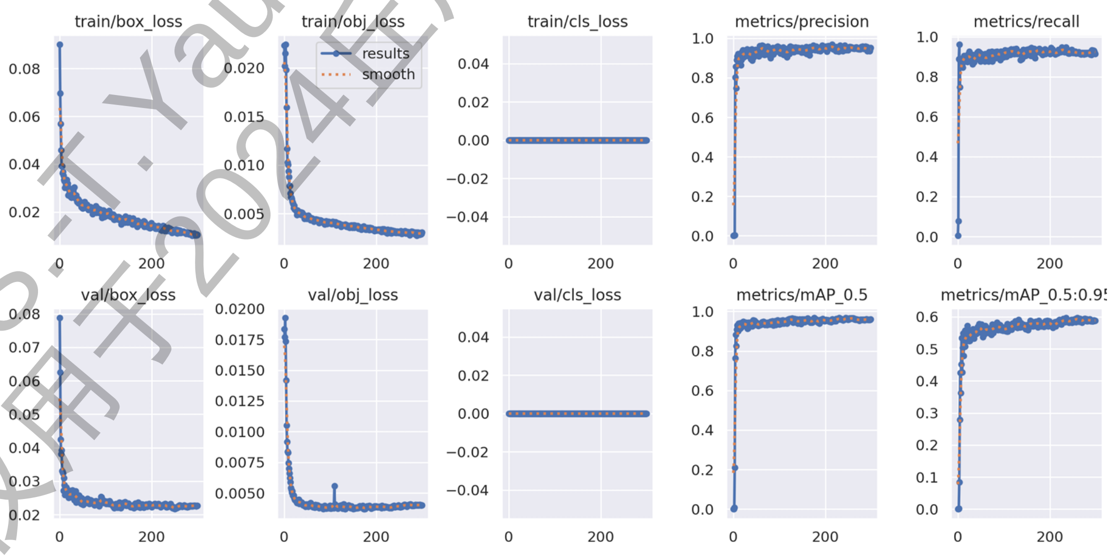


Figure 16. Training results of traffic light detection

It can be seen from figure 17 that loss gradually decreases during the training process for both

2024 S.-T. Yau High School Science Award
仅用于2024年丘成桐中学科学奖论文公示

box_loss and obj_loss, indicating that the model is continuously optimized. Also, train_loss and val_loss decrease gradually and converge to similar values. This shows that there is no obvious overfitting phenomenon in the training process of the model, and the performance of the model is consistent on the training set and the validation set. Both mAP@0.5 and mAP@0.5:0.95 show good performance. The high value of mAP@0.5:0.95, especially, indicates that the model is stable under different IoU thresholds and has strong generalization ability. This means that the model not only has good detection performance at looser IoU thresholds (such as 0.5), but also at more stringent thresholds (such as 0.95).

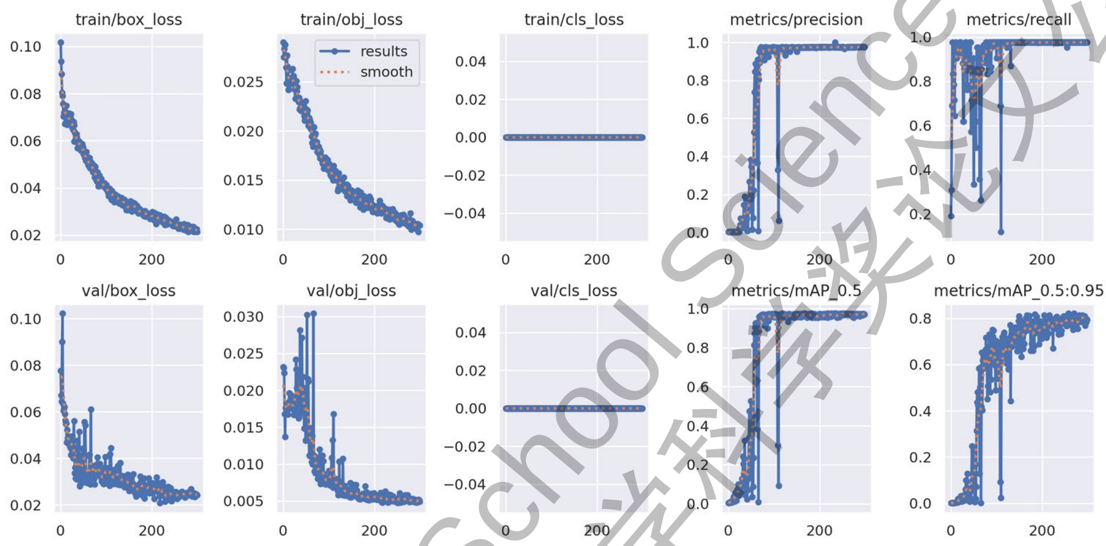


Figure 17. Training results of crosswalk detection

For figure 18(a), the model successfully identifies all the crosswalks and accurately frames the upper and lower boundaries of the crosswalks. This precise boundary framing not only enables the accurate detection of the crosswalks, but also facilitates the subsequent auxiliary traffic light selection algorithm to ensure that the detection can be conducted in the right area. Each detection is accompanied by a fairly high confidence level (0.9), demonstrating a high degree of trust in the identification results by the model. For image 18(b), except for the traffic light in one picture that is too small to be recognized, all the other small target traffic lights are effectively and accurately detected, and the detection results have high confidence (0.7-1.0). This shows that the model can provide reliable and accurate identification results even when dealing with small targets.

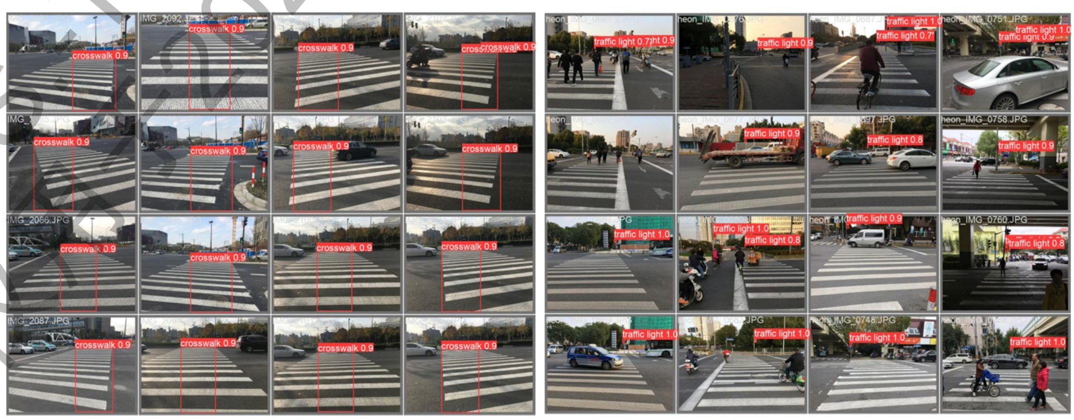
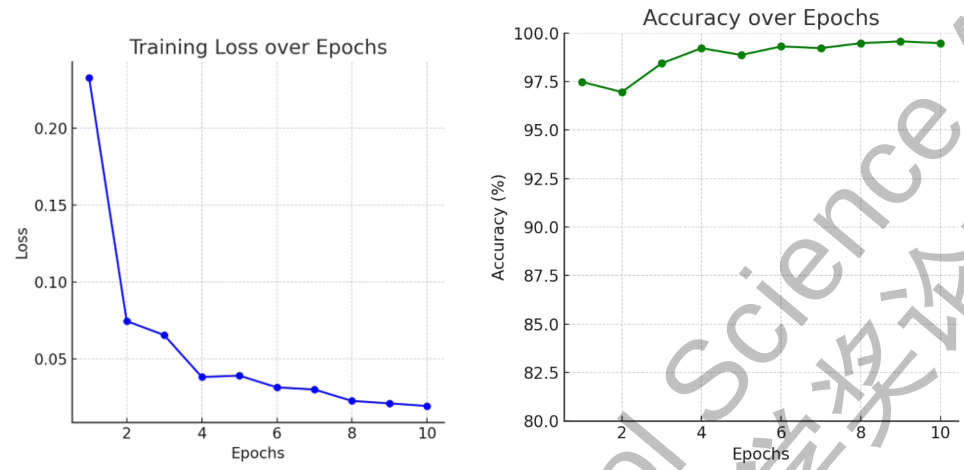


Figure 18. (a) Detection results for crosswalk (b) Detection results for traffic light

2024 S.-T. Yau High School Science Award 仅供学术交流公示

4.4 Performance of color detection

We use two-layer CNN to implement color detection. Figure 19 represents both training loss and accuracy with 10 epochs. The loss, shown in figure 19(a), decreases from approximately 0.23 to about 0.02, while the accuracy in the figure 19(b) increases to a peak of about 99.57%, with only slight fluctuations between epochs.



(a) (b)

Figure 19. Two-layer CNN train results for color detection

We perform a sensitivity analysis to evaluate the robustness of our traffic light classification model under different image conditions. Specifically, we explore how changes in brightness, contrast, and noise affect the performance of the model. We set a variation x to the six values with common difference respectively (as shown in table 3) and change the dataset with modified brightness, contrast, and noise based on the value of x separately. For brightness variation, we set

$$I_{new} = I_{orig} \times \text{factor}_{\text{brightness}} \tag{17}$$

where $\text{factor}_{\text{brightness}}$ is randomly chosen from $(1-x, 1+x)$, and for contrast variation, we set

$$I_{new} = (I_{orig} - 128) \times \text{factor}_{\text{contrast}} + 128 \tag{18}$$

where $\text{factor}_{\text{contrast}}$ is randomly chosen from $(1-x, 1+x)$. Finally, for noise variation, we set

$$I_{new} = I_{orig} + \text{noise} \tag{19}$$

where noise $\sim \mathcal{N}(0, \sigma)$, $\sigma = \text{factor}_{\text{noise}} \times 255$, $\mathcal{N}(0, \sigma)$ is a normal distribution with a mean of 0 and a variance of σ.

As presented in table 3, the sensitivity analysis shows that contrast is the most robust factor for CNN networks to distinguish between red and green lights at all contrast levels, and the model maintains very high accuracy, consistently above 98%, reaching a maximum of 99.65% with a contrast coefficient of 1.3. Brightness is the second robust factor, with accuracy consistently higher than 92%, reaching a peak of 99.57% at brightness coefficients of 0.9 and 1.1. While the network performs well at different brightness levels, it is slightly more sensitive to brightness variation than to contrast. Noise has a greater impact on performance. As the noise factor decreases, the accuracy increases from 54.73% at 0.6 to 99.48% at 0.1.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

Table 3. Sensitivity analysis

Brightness Factor	Brightness Accuracy (%)	Noise Factor	Noise Accuracy (%)	Contrast Factor	Contrast Accuracy (%)
0.5	92.71	0.1	99.48	0.5	98.27
0.7	99.13	0.2	98.96	0.7	99.13
0.9	99.57	0.3	90.55	0.9	99.57
1.1	99.48	0.4	77.36	1.1	99.57
1.3	99.22	0.5	65.13	1.3	99.65
1.5	99.05	0.6	54.73	1.5	99.57

4.5 System performance evaluation

The traditional YOLOv5 model detects and labels both vehicle traffic lights and crosswalk traffic lights in the image without making any distinctions, which results in difficulties in selecting the right crosswalk traffic light. Therefore, for comparison, we select the traffic light with the largest pixel size from multiple traffic lights detected by traditional YOLOv5, named YOLOv5+Pixel. We also change some parameters in our model to show how the accuracy changes.

Table 4 shows the simulation results. Score weight (α, β) are factors used to adjust the weights between the object's distance to the crosswalk and traffic light's area. Selected area (m, n) is the percentage of the selected area according to the crosswalk's bounding box. For YOLOv5+Pixel, the color detection is done by bounding box classification from YOLOv5. Our algorithm uses the designed two-layer CNN for color detection.

Table 4. Comparison of models with different strategies

Model Type	Score weight (α, β)	Selected area (m, n)	Color detection	Accuracy
1. YOLOv5+Pixel	N/A	Full image	YOLOv5 self-classifies	92.0%
2. Our Alogrithm	0.0, 1.0	Full image	Color detection module	96.8%
3. Our Alogrithm	0.5, 0.5	Full image	Color detection module	96.7%
4. Our Alogrithm	0.0, 1.0	50, 20	Color detection module	97.2%
5. Our Alogrithm	0.3, 0.7	50,20	Color detection module	97.5%
6. Our Alogrithm	0.5, 0.5	50,20	Color detection module	97.5%
7. Our Alogrithm	0.7, 0.3	50,20	Color detection module	96.3%
8. Our Alogrithm	1.0, 0.0	50,20	Color detection module	95.3%
9. Our Alogrithm	0.0, 1.0	30,20	Color detection module	97.0%
10. Our Alogrithm	0.3, 0.7	30,20	Color detection module	97.0%
11. Our Alogrithm	0.5, 0.5	30,20	Color detection module	97.2%
12. Our Alogrithm	0.7, 0.3	30,20	Color detection module	96.0%
13. Our Alogrithm	1.0, 0.0	30,20	Color detection module	95.2%

Simulation results show that our algorithm achieves higher performance. Comparing the results in table 4, YOLOv5 +Pixel achieves a relatively lower accuracy of 92.0%. In contrast, our algorithm's accuracy improves significantly to 97.5%. This shows that our algorithm can substantially enhance the model's overall accuracy by associating crosswalk traffic light with crosswalk, detecting small object and specific color. Model 1 and 2 both use only pixels (the area of light) to eliminate multi-light interference in the full image. The results show that our algorithm has increased the detection accuracy from 92% to 96.8% due to the use of the SE module and the color recognition module. These results also show that, when other conditions are the same, changing the score weight (α, β) will affect the final accuracy. When α and β are relatively balanced (such as 0.5, 0.5), the accuracy of the algorithm is the highest, reaching 97.5% for selected area (50,

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

20) and 97.2% for selected area (30, 20). When it is completely biased towards one weight, the accuracy will drop. This shows that our weighting approach successfully balances both factors, and also shows that both distance to crosswalk and area of traffic light are essential for accurate detection of crosswalk traffic lights. Similarly, when other conditions remain unchanged, the simulation results of only changing selected area (m, n) show that appropriately reducing the detection area has a higher accuracy than using the full image. However, further reducing it will lead to a decrease in the detection rate, such as the results in model (2, 4, 9) and (3, 6, 11). This shows that appropriately limiting the detection area can reduce noise and improve detection accuracy.

Table 5 highlights the superiority of our proposed algorithm (score weight (0.5, 0.5), selected area (50, 20)) compared with the YOLOv5+Pixel. Our algorithm shows notable increases in precision (98.3% vs. 92.8%) and accuracy (97.5% vs. 92.0%), indicating better overall detection reliability. Additionally, it reduces the number of incorrect detections significantly while maintaining a higher recall rate (99.2%). It is worth noting that among these metrics, precision focuses on how accurate the positive predictions are, that is, of all the instances predicted as positive, how many are actually positive. Precision is important when false positives carry a high cost, such as in traffic light detection, where incorrect predictions could lead to harm or accidents. The results show that our algorithm has significant advantages in metric of precision.

Table 5. Performance comparison of our algorithm and traditional YOLOv5+Pixel

Model	Precision	Recall	Accuracy
Our Alogrithm	98.3%	99.2%	97.5%
YOLOv5+Pixel	92.8%	99.1%	92.0%

5. Conclusion

The paper presents a novel crosswalk traffic light detection algorithm designed to resolve the specific challenges faced by the visually impaired when crossing street. This work, as far as we know, is the most comprehensive work for the detection of crosswalk traffic light. In this work, we analyze the data characteristics to get the factors that the algorithm needs to have, so that we can construct the algorithm based on this priori knowledge, which effectively improves the accuracy of the algorithm. The proposed algorithm effectively resolves multi-light interference by associating traffic lights with the corresponding crosswalk, significantly improving detection accuracy. SE and CBAM are used to enhance the detection of small objects and crosswalks under varying conditions, respectively. A two-layer CNN is designed for detecting color of different types of crosswalk traffic lights. Experimental results demonstrate that our algorithm performs better than traditional YOLOv5 algorithm combined with maximum pixel detection, especially in distinguishing crosswalk traffic lights from other traffic lights, which effectively improves the travel safety of the visually impaired.

Accurate detection is crucial for safety. Although our algorithm has achieved better results, it is still not 100% safe. In the future, we need to optimize our algorithm modules, further analyze the images with detection errors to improve our algorithm, and expand the dataset to include more diverse crosswalk and traffic light configurations.

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

References:

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[2] X. Zhang, Y. Li, J. Wang, and Q. Liu, "Toward Effective Traffic Sign Detection via Two-Stage Fusion Neural Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1234-1245, Mar. 2023.

[3] R. Girshick, J. Donahue, T. Darrell, et al., "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 2015.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016.

[5] T.-Y. Lin, P. Goyal, et al., "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 2020.

[6] Q. Wang, Q. Zhang, X. Liang, Y. Wang, C. Zhou, and V. I. Mikulovich, "Traffic Lights Detection and Recognition Method Based on the Improved YOLOv4 Algorithm," *Sensors*, vol. 22, p. 200, 2022.

[7] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.

[8] S. Yan, X. Liu, W. Qian, and Q. Chen, "An end-to-end traffic light detection algorithm based on deep learning," in *Proc. 2021 Int. Conf. Security, Pattern Anal., Cybern. (SPAC)*, Chengdu, China, Jun. 18–20, 2021, pp. 370–373.

[9] https://github.com/ultralytics/yolov5 .

[10] W. Liu, Z. Wang, B. Zhou, S. Yang, and Z. Gong, "Real-time Signal Light Detection based on Yolov5 for Railway," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 769, no. 3, 042069, 2021.

[11] S. S. Krishna, V. Parisha, B. U. K. Varma, and C. Srinivasulu, "Traffic Light Detection and Recognition using YOLOv8," in *Proc. 2024 3rd Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, 2024.

[12] https://yolov8.com/ .

[13] P. S. Swami and P. Futane, "Traffic Light Detection System for Low Vision or Visually Impaired Person Through Voice," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-5.

[14] S. Eto, Y. Wada, and C. Wada, "Convolutional Neural Network Based Zebra Crossing and Pedestrian Traffic Light Recognition," *Journal of Mechanical and Electrical Intelligent System*, vol. 6, no. 3, pp. 1-11, 2023.

[15] V. Rao and H. Nguyen, "A computer vision based system to make street crossings safer for the visually impaired," *Journal of High School Science*, vol. 8, no. 2, pp. 253-266, 2024.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2018.

[17] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sept. 2018.

[18] https://www.makesense.ai/ .

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

致谢

与这篇论文相关的故事最早可以追溯到 2022 年 11 月，我们在珠江路上偶遇了一位盲人老爷爷。繁华的珠江路，车来人往，老爷爷拄着盲杖，但是由于盲道被电瓶车占用，老爷爷走得艰难而危险，让我们很难过。从那时候起，我们就开始调研南京盲人出行的现状。

本文首先要感谢的是调研过程中给我们提供帮助的老师和朋友，特别是一些视觉障碍朋友。论文工作的初步想法源于去盲校的一次交流。半年前，在南京市盲人学校老师的带领下，我们得以和该校多名学生进行交流。交流过程之中，一位同学说他们出门时过马路非常不方便。他不是盲人，是视弱者，他们无法清晰看到马路对面人行横道红绿灯的状态，安全地通过斑马线。在如今这个汽车都有了智能红绿灯信号识别系统的时代，他们却依旧面临着如何安全通过斑马线的难题。在这一年多和盲人朋友的接触中，我们了解到与我们平时想的不同，实际上通过语音辅助等，他们大部分能使用智能手机，而且用得还不错，照相这些功能他们基本都能使用，甚至还会拍照发朋友圈。于是，我们决定设计结合手机拍照，帮助视弱群体安全过马路的交通灯识别算法。我们一开始尝试用已有的机器学习算法来解决这个问题。然而，在实际操作中我们发现该方案并不能满足我们排除干扰红绿灯，在多个红绿灯中选择出正确的人行横道红绿灯的要求。通过对图片的分析，我们发现人行横道红绿灯总是与人行横道共同存在，于是我们利用人行横道作为我们判别正确红绿灯的参照，最终达到了更高的检测率。在这一过程中，同学、朋友、还有在各地的叔叔阿姨们为我们拍摄了来自不同城市的人行横道红绿灯图片，极大地丰富了数据的多样性，对此，我们深表感谢。

我们也衷心地感谢东南大学张川教授为我们项目所提供的一切帮助，张老师同时也是张景涵同学英才计划的导师。从发现问题到完成算法，我们中途遇到过众多难题。无论是大到方案设计，还是小到程序上的报错，张老师总是无偿地给予我们引导性的指导。相较于直接告诉答案，张老师教给我们的是解决问题的方法和严谨的学术思维，让我们受益匪浅。同时，张老师开放服务器给我们使用，大大的缩短了我们程序运行的时间，显著提升了我们的效率。

作为队友，我们配合默契，分工明确。张景涵同学主要负责方案设计、测试，田桢干同学负责模块设计、编程和测试，我们共同拍摄、收集和处理了实验所需要的图片，并撰写了论文，论文第一节、第二节和第三节部分和结论主要由张景涵同学撰写，论文第三节部分和第四节主要由田桢干同学撰写，数据分析由我们共同完成。在整个过程中，我们遇到了不少困难，如我们一开始设计的算法精度并不高。为了解决这个问题，我们对错误图片逐一分析，总结图片特点。根据分析结果，改进算法，调整参数，进一步排除了干扰目标，提高识别率。同时，我们还多次实验，摸索人行横道红绿灯在照片中呈现的特征，并在算法中体现，终于得到了满意的结果。后继我们还将进一步优化，争取获得更好的性能，这是一个非常让人着迷的过程。在此期间，要感谢我们彼此一直以来的理解和配合，特别是冲刺阶段互相的鼓励和奋战，令人难忘。

我们还要感谢互联网提供的各种资源，感谢 CSDN 社区、GitHub 上大家的无私分享，作为受益者，我们也将代码上传，欢迎有兴趣的同学下载交流:

https://github.com/TmFnaXNh/crosswalk-traffic-light-detection/tree/main。

感谢母校南京外国语学校给了我们一个宽松的学习环境，在南外的四年，让我们能够自由探索，也让我们能结识更多优秀的老师和同学。

最后，感谢丘成桐中学科学奖给了我们这个成长的机会。作为一名高中生，我们得以在上大学之前就较为系统地了解学科知识，实验方法，论文写作。这不仅仅帮助我们掌握了知识与技术，也坚定了我们在这条路上继续钻研下去的决心。

2024 S.T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示

参赛学生和指导教师简历:

张景涵: 南京外国语学校高二年级学生，入选全国中学生英才计划和光启青年学者项目、在美国数学竞赛（AMC10）中取得全球前 1%的成绩、入选 USAMO、在英国高级数学挑战赛（SMC）中获金奖、在美国高中数学建模竞赛(HiMCM)中获 M 奖、在第二十五届世界哲学大会独立发表论文“行为主义人工智能过度泛化的哲学之思”并作为该届最年轻参会者做现场报告、在剑桥 Rethink 写作比赛获得 Honorable Mention、获全国青少年科技创新大赛江苏省一等奖、21 世纪杯全国英语演讲比赛江苏省一等奖等。

田桢干: 南京外国语学校高二年级学生，在美国计算机奥林匹克竞赛(USACO)中进入铂金组、在美国数学邀请赛(AIME)中获得 12/15 并获得资格进入决赛、在高中数学建模竞赛(HiMCM)中排名全球前 5%、在美国数学竞赛(AMC 10)中获得 120/150，全球排名前 5%、在青少年信息学奥林匹克竞赛提高组(CSP-S)中获得一等奖、在全国青少年信息学奥林匹克联赛(NOIP)中获得一等奖、参加普林斯顿计算机算法夏校(PACT 2023)。

张川: 东南大学教授、博士生导师，东南大学青年首席教授，紫金山实验室课题联合负责人。获国家自然科学基金重点项目、优青项目，强国青年科学家，中国工程院“中国工程前沿杰出青年学者”，江苏青年五四奖章，江苏省杰出青年基金，江苏省“六大人才高峰”高层次人才 A 类，江苏省 333 高层次人才，中学生英才计划导师等。获国家级教学成果奖二等奖、江苏省教学成果奖一等奖、东南大学“杰出教学奖”、东南大学我最喜爱的研究生导师“十佳导师”等。担任 IEEE 电路与系统学会杰出讲师、IEEE J. Emer. Top. Circuits Syst.高级编委会委员、IEEE 电路与系统学会 CASCOM 技术委员会主席、IEEE Trans. on Mobile Comp.指导委员会委员、及 IEEE 电路与系统学会旗舰会议 IEEE ISCAS 2021 Tutorial Speaker 等。担任 IEEE Trans. Circuits Syst. II 副主编、IEEE Trans. Signal Process.副主编、IEEE Open J. of Circuits and Syst.副主编、IEEE J. Emer. Top. Circuits Syst.客座主编等。

2024 S.-T. Yau High School Science Award
仅用于2024丘成桐中学科学奖论文公示