Research on Optimization of Belt Conveyor Foreign Object Detection System Based on Improved YOLOv5 Algorithm and Deep Learning Enhancement

Wang Jin[1], Zhang Yan[1],Dou Litong[1], Zheng Mingliang[1]*, Ronaldo A Juanatas[2]

[1]School of Mechanical and Electrical Engineering, Huainan Normal University, Huainan 232038,China

[2]University of Technology of the Philippines, Manila, Philippines

*E-mail: liangmingzheng@hnnu.edu.cn

Abstract: This study aims to design a belt conveyor foreign object recognition system based on the Easy Language programming environment and the YOLOv5 deep learning model, with the goal of enhancing safety and efficiency on the production line. The overall architecture of the system encompasses core modules such as data collection, image preprocessing, model training, and result output. The system employs Easy Language as its programming language and integrates the YOLOv5 object detection algorithm. Through real-time video monitoring, it precisely identifies foreign objects and belt damage during the transportation process of the belt conveyor, effectively preventing and reducing accidents and losses on the production line. Experiments have demonstrated that the system significantly enhances the accuracy and efficiency of foreign object recognition in continuous conveyor belt operations.

Keywords:easy language; YOLOv5;Belt conveyor; Foreign object identification; Belt damage

INTRODUCTION

In the field of industrial manufacturing, as the core handling tool, the safety, continuity and efficiency of conveyor belt system are critical to the entire production process. However, the complex and changeable working environment often causes the conveyor belt to face challenges such as invasion of foreign objects, belt damage and operation deviation, which seriously affects the production efficiency, and even causes equipment damage and safety accidents [1]. The traditional manual inspection method is inefficient and has high false detection rate, which is difficult to meet the high requirements of modern industry for safety and efficiency. Therefore, it is urgent to develop a system for automatic detection of foreign matters in the conveyor belt. The foreign matter identification system of belt conveyor based on Easy Language and YOLOv5 proposed in this research uses deep learning algorithm and easy language programming to realize real-time monitoring and accurate foreign matter identification of belt conveyor [2]. The system can not only significantly improve the detection efficiency and accuracy, reduce production costs and maintenance costs, but also effectively prevent equipment damage and safety accidents caused by foreign object intrusion, and improve production safety. In addition, the system is easy to deploy and maintain, which promotes the development of industrial automation technology and has broad application prospects and important theoretical and practical significance.

Object detection is a key challenge of machine vision, which aims to identify the types and positions of objects in images. Traditional methods include three stages: target area determination, feature extraction and classification. However, they are limited by artificially defined features and are difficult to deal with complex scenes, lacking robustness. The introduction of deep learning breaks this bottleneck, simplifies the learning process and enhances adaptability by mining the deep characteristics of data. Therefore, deep learning technologies such as YOLOv5 have gradually replaced traditional methods and become the mainstream in the field of object recognition [3]. This paper proposes a belt conveyor foreign object recognition system based on Easy Language and YOLOv5, aiming at improving the belt conveyor automation and production efficiency. The system solves the problems of low efficiency and high false detection rate of traditional manual inspection methods, effectively improves the detection efficiency and accuracy through real-time monitoring, and reduces production costs and maintenance costs. In addition, the system helps to prevent equipment damage and safety accidents and improve production safety. The combination of Easy Language and YOLOv5 not only improves the application value of the system, but also promotes technological innovation and development [4].

OBJECTIVES

System theoretical basis

2.1 Principle of YOOv5

YOLO algorithm performs target detection through convolutional neural network and assigns category labels to targets. In terms of efficiency and speed, YOLOv5 can simultaneously achieve rapid detection and high accuracy. YOLOv5 architecture is designed based on the evolution of YOLOv3 [5]. The overall model structure of YOLOv5 is shown in Figure 1.

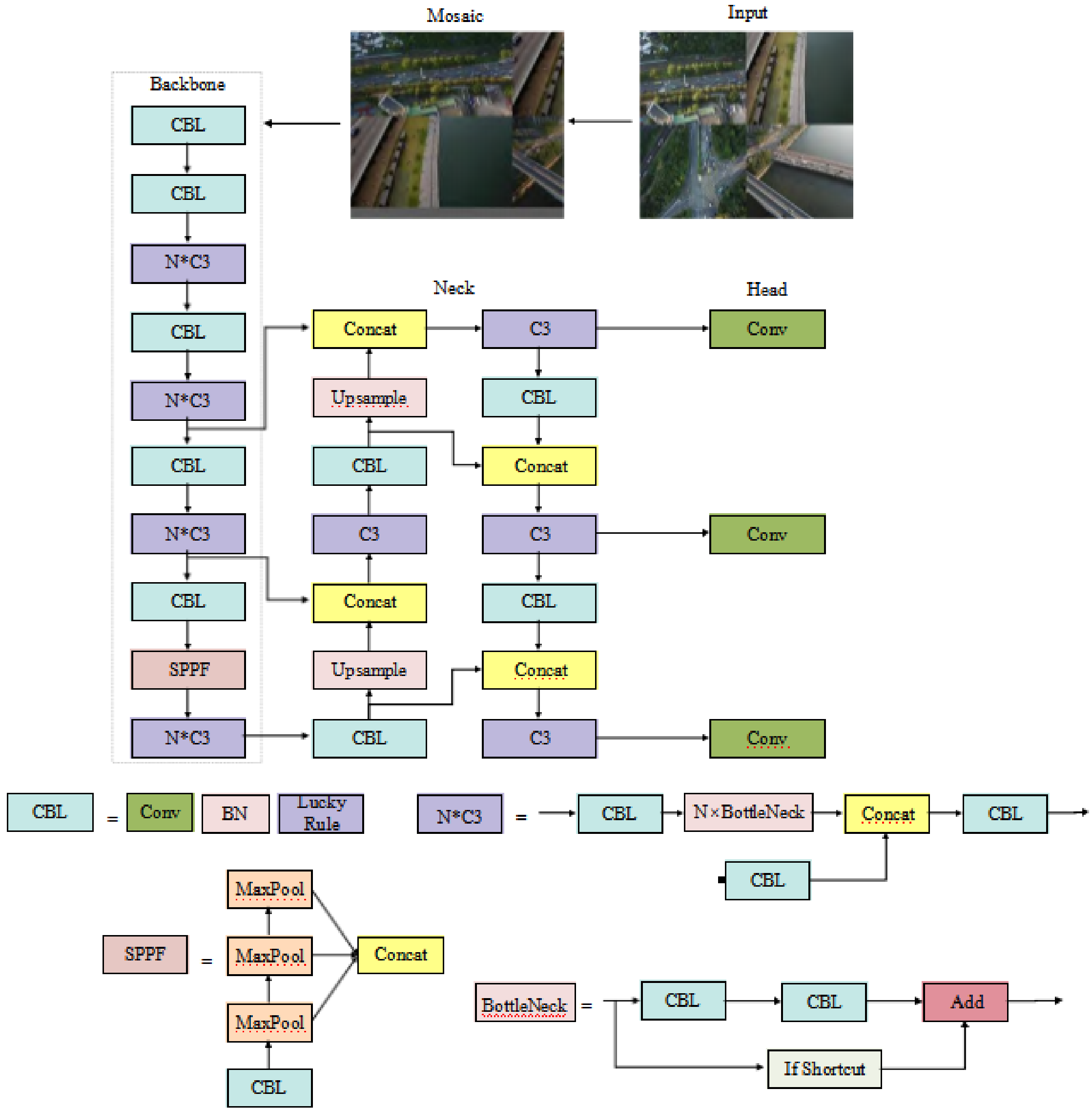


Figure 1 Overall Model Structure of YOLOv5

There are four different models in YOLOv5, namely, Yolov5s, Yolov5m, Yolov5l and Yolov5x [6]. The performance comparison curve between the four YOLOv5 models and the EfficientDet detection algorithm is shown in Figure 2. The abscissa of the graph represents the reasoning time of the model on the GPU. Ideally, the smaller the value, the better; The ordinate represents the average accuracy (AP) of the model when testing the COCO dataset. The higher the index, the better . The analysis results show that Yolov5s version achieves faster reasoning speed while maintaining a higher AP value.

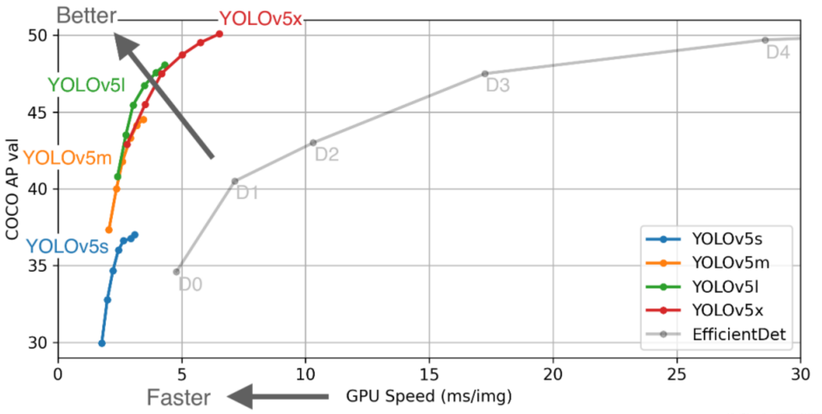


Figure 2 Performance display of YOLOv5 models

Yolov5s is the model version with the smallest depth and width in YOOv5 series. Its model structure is shown in Figure 3.

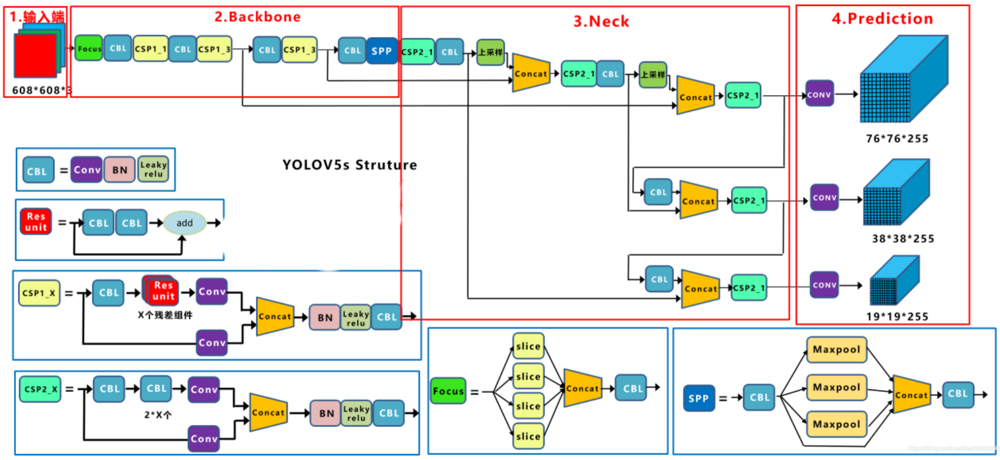


Figure 3 YOLOv5s Model Structure

YOLOv5 model mainly consists of three parts, namely Backbone, Neck and Head, as shown in Figure 4. Backbone is mainly responsible for feature extraction of input pictures. Neck is responsible for multi-scale feature fusion and transferring these features to the prediction layer. Head made the final regression prediction [7].

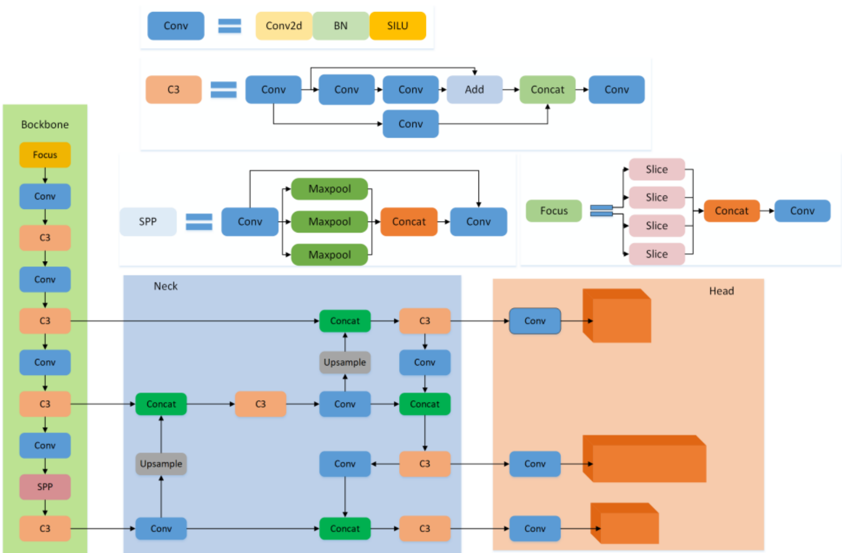


Figure 4 Backbone, Neck and Head Structure Diagram

METHODS

Method of calling YOLO model in easy language

The combination of Yi language and YOLO model to realize the automatic recognition of foreign matters in the belt conveyor can be realized mainly by the following two methods:

Method 1: Call YOLO model through Python script as intermediary

Preliminary preparation: First, prepare the pre training weight file and configuration file of YOLO model, which can usually be obtained from official sources or trusted channels. At the same time, you need to install the Python environment and configure necessary deep learning libraries, such as PyTorch or ONNX Runtime, to support the loading and reasoning of YOLO models [8].

Python scripting: Write a script in Python to load YOLO model and detect the target of the input belt conveyor image. The script should be able to receive image input and return the detected foreign object information.

Easy language calls Python script: In easy language, the above Python script is called by executing external programs. Easy language program can be responsible for image acquisition and pre-processing, and then transfer the processed image to Python script for target detection. Finally, the Yi language program receives the detection results returned by the Python script and performs subsequent processing or display [9].

Method 2: directly call the API of YOLO model (technical conversion is required)

Model transformation and encapsulation: convert YOLO model from original format (such as PyTorch. pt file) to a format that can be directly called by languages. This usually involves converting the model to ONNX format, and then using C++or other programming languages for encapsulation to create API interfaces that can be easily called by languages. This process requires deep experience in programming and deep learning model transformation [10].

Easy language calls API: In easy language, you can directly call YOLO model by calling the above encapsulated API interface. Easy language program can be responsible for image acquisition, pre-processing and model call, as well as the reception and processing of detection results.

3 System overall design

This paper focuses on capturing the running image of the belt conveyor in real time through the high-definition camera, and implementing the annotation and pre-processing process for the image, covering the noise removal, enhancement, segmentation and other links of image quality improvement, as well as the technology to enhance the accuracy of foreign object identification [11]. YOLOv5 algorithm is applied to identify the preprocessed image, extract the features of foreign matters on the belt conveyor, and then train and optimize the model to ensure that the model can accurately identify foreign matters [12]. Using easy language program, the model is deployed to identify foreign matters on the belt conveyor and output the identification results. Analyzing these data can reveal the running state and development trend of belt conveyor, and provide important reference for the formulation of preventive maintenance and optimization strategy of equipment.

3.1 Overall architecture of monitoring system

Determine the overall structure of the system according to the functional requirements of the monitoring system, as shown in Figure 5.

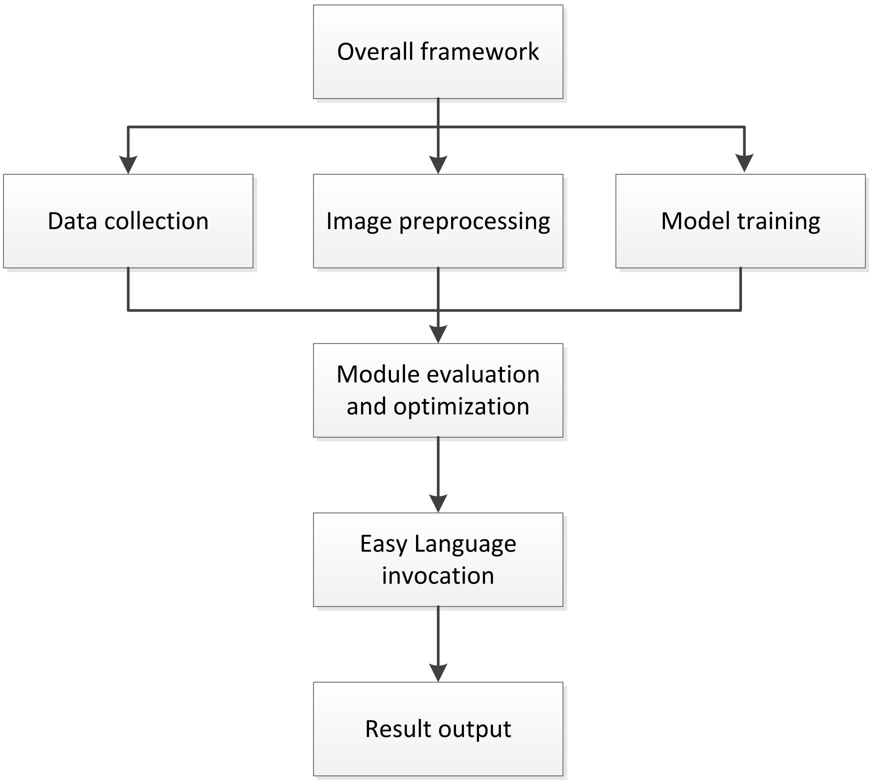


Figure 5 Overall System Framework Diagram

Data acquisition and pre-processing module: This module focuses on collecting image data from the belt conveyor and performing a series of pre-processing steps, such as image size adjustment, color enhancement, etc., in order to lay a solid foundation for subsequent target detection tasks, so as to improve the accuracy of recognition.

Model training module: conduct in-depth training on YOLOv5 model with the collected image data set. The module transmits the input image to the model, and detects foreign objects in the image through the reasoning ability of the model.

Model evaluation and optimization: use the data set to evaluate the trained model, calculate the accuracy index of the model on the target detection task, and refine and adjust according to the actual needs to achieve a good training effect [13].

Easy language interface and result output: With the interface provided by Easy language, the system can easily call the trained YOLOv5 model and transfer the preprocessed image data. After the model runs, the detected foreign matter information will be returned, which will be displayed to the operator through the system interface for observation and analysis.

The flow chart is shown in Figure 6 below, which depicts the whole process from starting the system to ending the operation. The system process is briefly described as follows:

(1) Start the system: first, start the belt conveyor foreign object recognition system.

(2) Image acquisition: acquire image data from the belt conveyor in real time through a camera or other image acquisition equipment.

(3) Preprocessing image: preprocess the acquired image, including size adjustment, color space conversion, image enhancement, etc., to improve the accuracy of target detection.

(4) Model call and detection: use the easy language interface to call YOLOv5 model, and input the preprocessed image data into the model for foreign object detection.

(5) Result display: The foreign matter information detected by the model is displayed through the system interface, providing intuitive visual feedback for the operator.

(6) End of operation: the system will end operation after all detection tasks are completed.

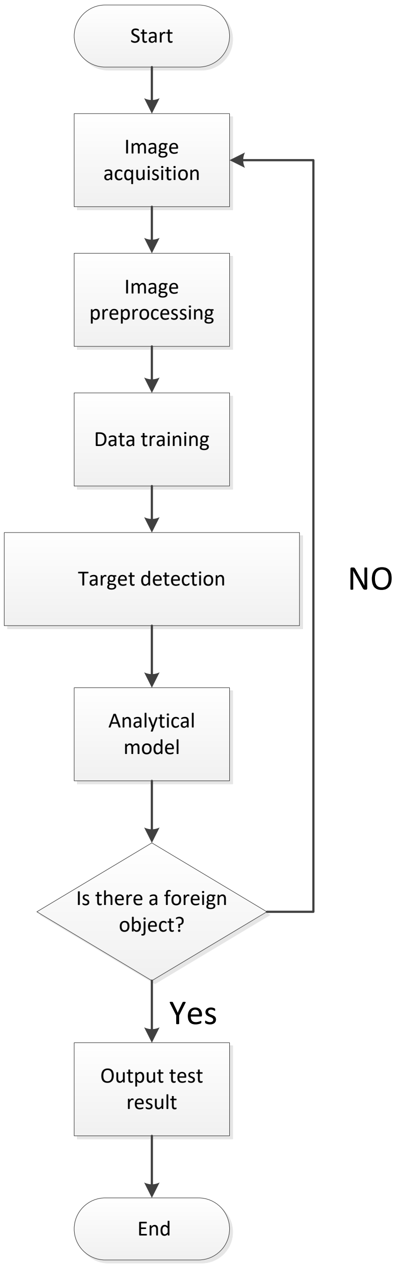


Figure 6 System flowchart

In the practical process of architecture design, special attention is paid to the integration and innovation of technology. The combination of Yi language and YOLOv5 provides a convenient and efficient development environment for the system. Yi language, with its intuitive and easy to understand programming paradigm and rich functional components, greatly simplifies the development process of the system and reduces development difficulty. YOLOv5, with its excellent object detection algorithm and efficient real-time processing capabilities, endows the system with powerful foreign object recognition capabilities. The deep integration of the two not only improves the development efficiency of the system, but also significantly enhances the accuracy and efficiency of foreign object recognition while maintaining system flexibility and stability.

3.2 Data Collection and Sample Annotation

It is crucial to build a comprehensive and well labeled image dataset when developing a belt conveyor foreign object detection system. This involves determining annotation categories, collecting image data, classifying and filtering to ensure quality, naming storage systems for easy retrieval, and preparing annotation schemes and proficiently using annotation tools to ensure accuracy and efficiency. These meticulous preparations have laid a solid foundation for model training.

3.2.1 Data Collection

As shown in Figure 7, the image acquisition tool can choose from three modes: full screen, window, and custom area.

Figure 7 Image acquisition tool

Normalize the processed image data, usually by scaling the pixel values to a range of 0-1, in order to maintain consistency with the input data during model training. Collect image data as shown in Figure 8.

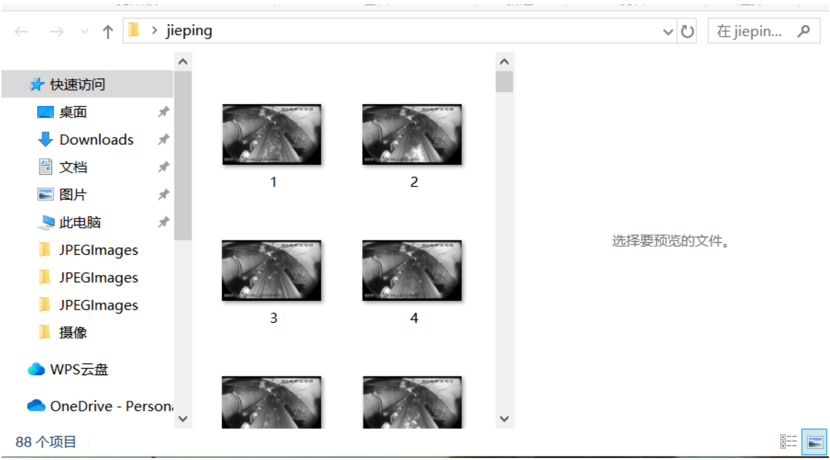


Figure 8 Collecting Image Data

3.2.2 Annotated Samples

Experimental implementation often requires a large amount of image resources, and its annotation work is quite time-consuming. Automated image recognition technology has become a feasible choice to improve annotation efficiency. The basic processing flow includes: using technology to preliminarily identify image content, followed by manual review and correction, and ultimately achieving efficient and accurate labeling.

Data preparation stage: Firstly, it is necessary to build a large image database containing multiple different elements. Subsequently, these images need to be carefully classified, distinguishing between images containing and not containing foreign objects, and forming a clear and distinguishable set of positive and negative samples [14]. This type of classification work is crucial for subsequent automated processing as it provides the necessary foundational data support.

Data preprocessing: It is crucial to perform a series of preprocessing steps before machine learning or deep learning analysis on images, as shown in Figure 9. The purpose is to improve the efficiency and accuracy of the model in processing these images. These adjustments help ensure that the model can extract information more effectively from the image, resulting in more accurate analysis results.

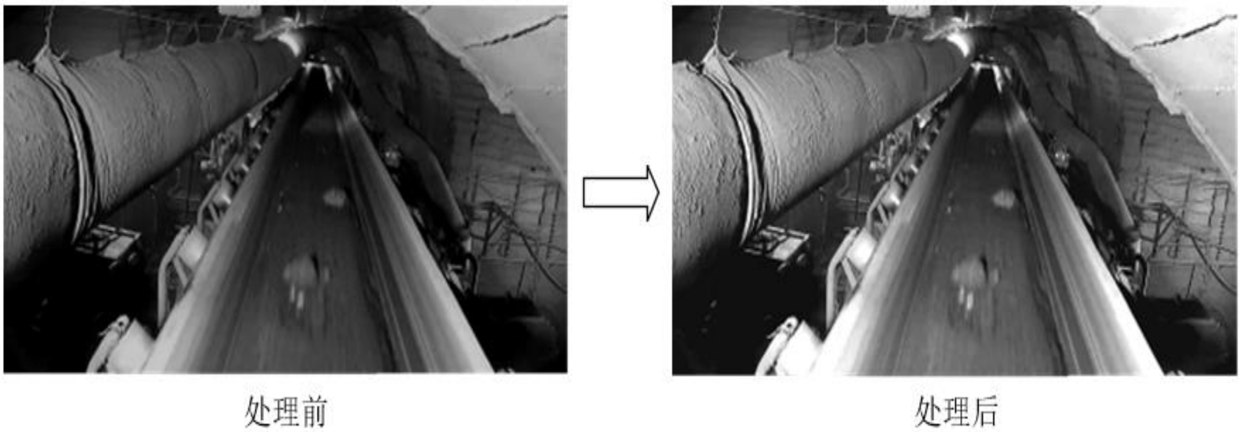


Figure 9 Image Enhancement Processing

Before performing automatic recognition, it is important to ensure that the dataset is large enough and representative, and that it is labeled correctly. When automatically annotating image samples, if there are any unreasonable annotations, they can be manually adjusted. To ensure accurate labeling of all samples, foreign object labeling is shown in Figure 10, and belt damage labeling is shown in Figure 11.

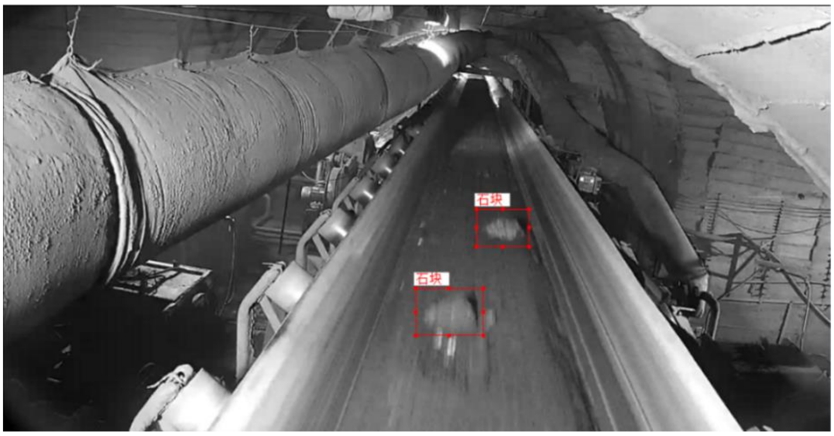


Figure 10 Foreign Object Labeling

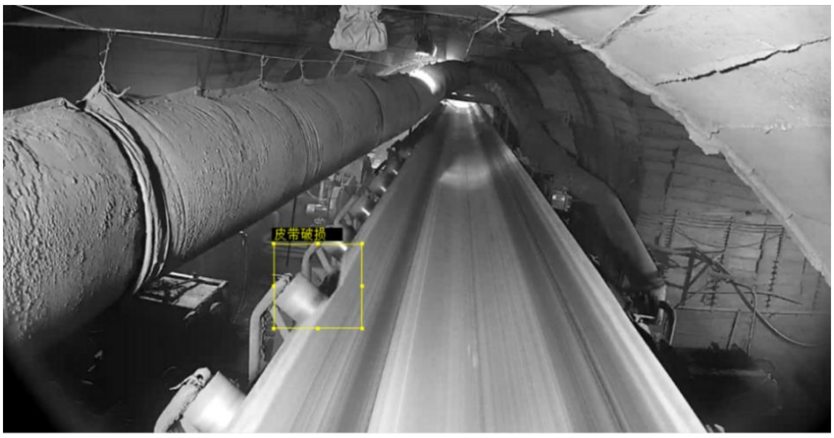


Figure 11 Belt Damage Labeling

3.3 Model Training and Recognition System Design

3.3.1 Model Training Parameters

Model training software is a tool used to train machine learning and deep learning models, providing rich functionality and interfaces to assist developers in tasks such as data processing, model construction, training, and evaluation. The training parameters of the YOLO model are shown in Figure 12, and can be adjusted according to experimental requirements.

Figure 12 Training Parameters

Here are some YOLO model training parameters and their meanings:

(1) Batch size: This parameter defines the number of samples to be processed in each training iteration.

(2) Learning rate is one of the key factors controlling how parameters are adjusted during model training. Each iteration will affect the speed of parameter adjustment [15].

(3) Epochs: In machine learning, it refers to the number of times a model processes the entire training dataset. It directly affects the learning efficiency and ultimate maturity of the model's performance.

(4) Width (input image width): The image width setting is used to specify the horizontal size of the input image.

(5) Height (input image height): This parameter defines the vertical size requirement for the input image of the model. The YOLO model typically uses a square image shape, with common sizes including 416x416 or 608x608. Choosing such dimensions helps the model effectively identify and locate target objects of different sizes, as shown in Figure 13.

（6）subdivisions： It refers to the number of partitions per batch during forward and backward propagation.

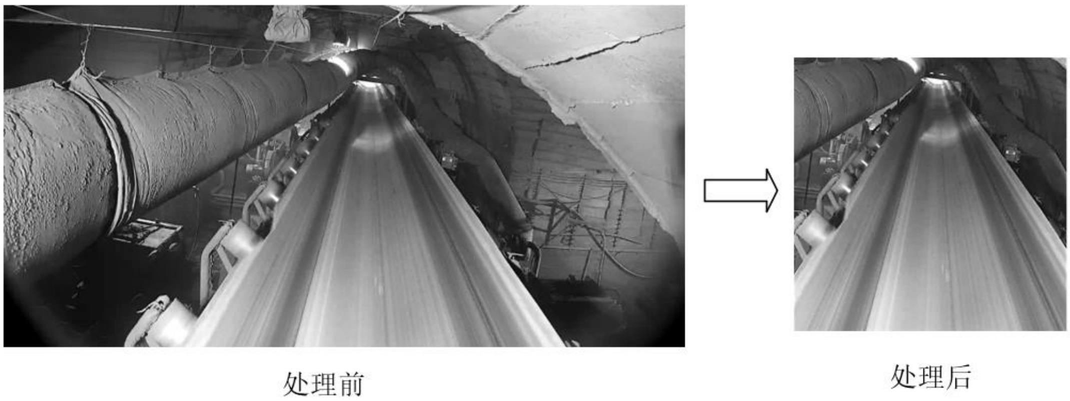


Figure 13 Size processing of images

3.4.2 Model Training

In the training process of the model, the loss function and iteration are two important parameters [16]. The current loss value is on the y-axis, and the number of iterations is on the x-axis.

Set the model training parameters, such as batch size set to 64, learning rate set to 0.001, input image width and height set to 416, etc. After the initial parameter settings are completed, use the prepared dataset and configured parameters to start training the model. If the computer is training for the first time, the training tool will automatically configure the environment in the background, which may take a few minutes. After the configuration is completed, the model will start training. As shown in the training curve of the model in Figure 14, the training process of the model can be adjusted and analyzed based on its shape and trend:

Smooth descent: In the early stages of model training, the loss value is usually relatively high. As the number of training iterations increases, this value gradually decreases, indicating that the model is gradually adapting to and learning the core features of the data, which is a positive sign of improving model performance [17].

Convergence: When the loss value begins to stabilize and remains at a low level, it indicates that the model is approaching the so-called convergence state. At this stage, the improvement of the model will become smaller and the optimization space will become limited.

Severe fluctuations: During the training process, there may be issues with significant fluctuations in loss values, which may be caused by overfitting or excessively high learning rates [18].

High loss value: If the loss value remains too high and shows no signs of decreasing, this usually reflects several potential issues, such as model design being too simple, improper data processing, or incorrect parameter configuration during training. To improve model performance, it is possible to consider optimizing the model architecture and increasing the training period.

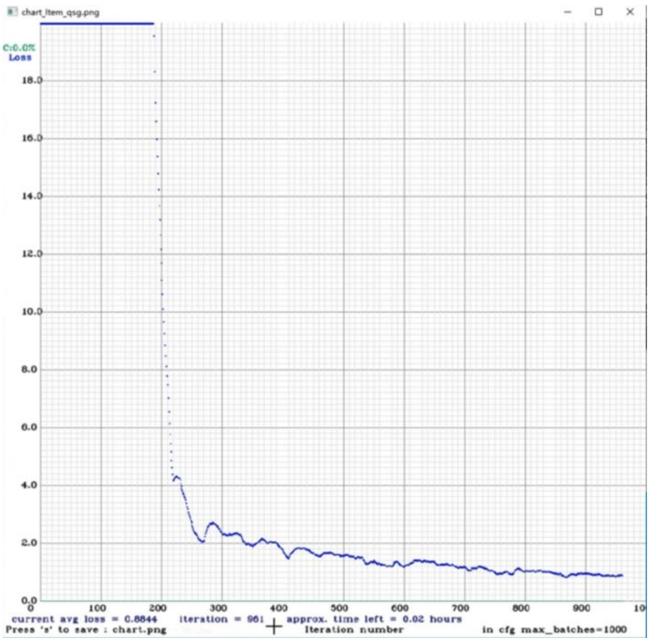


Figure 14 Model Training Curve

During the process of model training, key indicators such as loss value and iteration count are closely tracked and recorded to comprehensively evaluate the training progress and model effectiveness. The ideal situation is for the loss function value to approach zero, and the lower its value, the higher the recognition accuracy of the model. This training is set to a cycle of one hundred iterations, ensuring that at least 100 iterations are completed before generating the model file. Ultimately, the trained model will be extracted and properly stored.

3.4.3 Identification System Design

Design a foreign object recognition system by combining the YOLOv5 model with the Easy Language platform. Firstly, clarify the functional requirements and construct an operational interface framework; Integrate the model again to achieve efficient recognition. As shown in Figure 15.

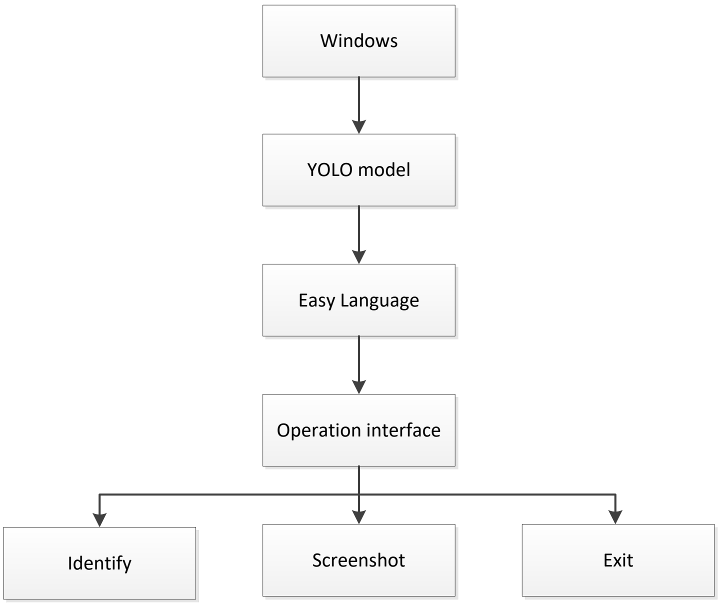


Figure 15 Identification System Interface Design Diagram

Open Easy Language, create a new project as shown in Figure 16, and a project creation interface will pop up. Select "Windows Window Program" and click "OK" to enter the Easy Language design interface. After entering the Easy Language design interface, use the built-in interface editing tools and components of Easy Language to create and layout the interface, as shown in Figure 17.

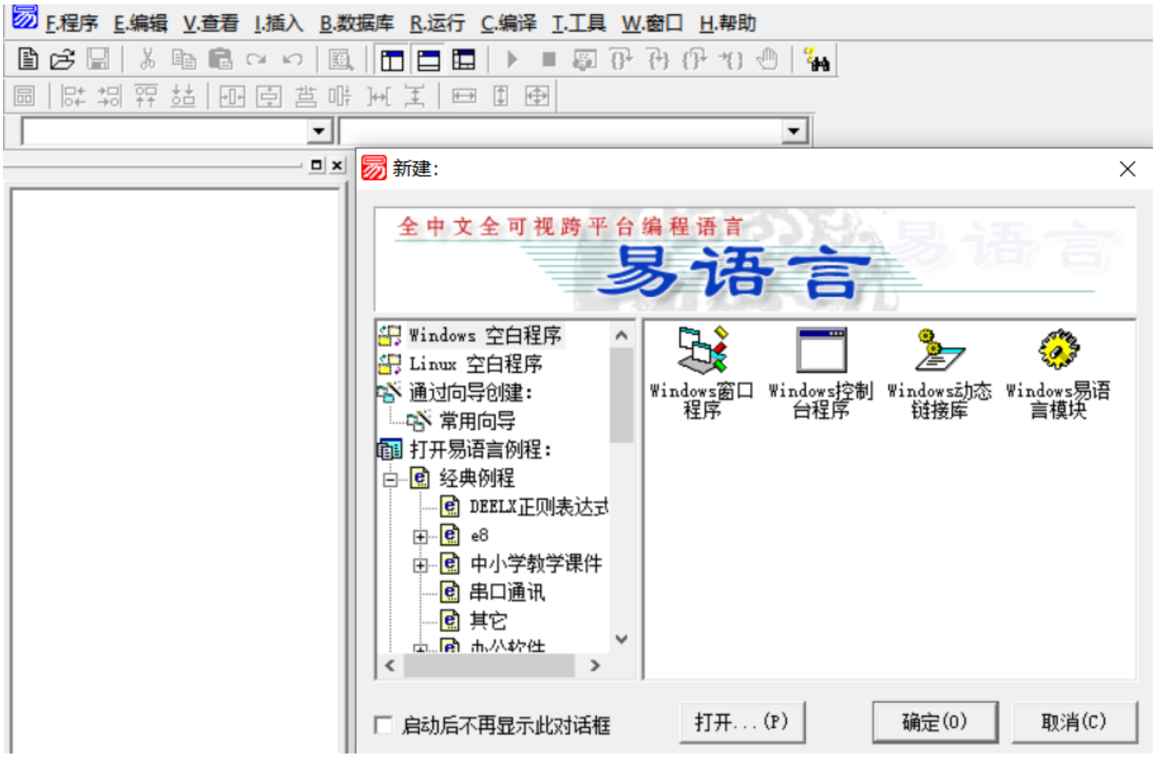
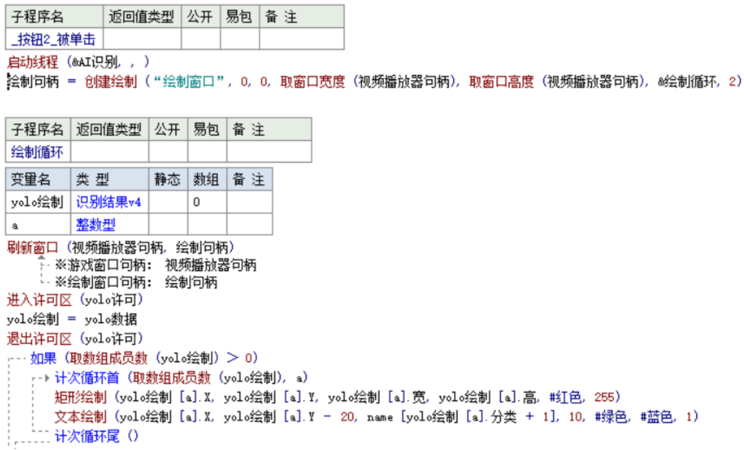


Figure 16: Creating a Project

Figure 17 System Interface Design

Firstly, after the user clicks the "Identify" button, the system receives an instruction to start the foreign object recognition program based on YOLOv5, as shown in Figure 18. Before running the recognition program, users can change some key recognition parameters that directly affect the accuracy and efficiency of recognition.



子程序名	返回值类型	公开	易包	备 注
_按钮2_被单击				

```
启动线程 (&AI识别, , )
绘制句柄 = 创建绘制 ("绘制窗口", 0, 0, 取窗口宽度 (视频播放器句柄), 取窗口高度 (视频播放器句柄), &绘制循环, 2)
```

子程序名	返回值类型	公开	易包	备 注
绘制循环				

变量名	类 型	静态	数组	备 注
yolo绘制	识别结果v4		0	
a	整数型			

```
刷新窗口 (视频播放器句柄, 绘制句柄)
    ※游戏窗口句柄: 视频播放器句柄
    ※绘制窗口句柄: 绘制句柄
进入许可区 (yolo许可)
yolo绘制 = yolo数据
退出许可区 (yolo许可)
如果 (取数组成员数 (yolo绘制) > 0)
    计次循环首 (取数组成员数 (yolo绘制), a)
        矩形绘制 (yolo绘制 [a].X, yolo绘制 [a].Y, yolo绘制 [a].宽, yolo绘制 [a].高, #红色, 255)
        文本绘制 (yolo绘制 [a].X, yolo绘制 [a].Y - 20, name [yolo绘制 [a].分类 + 1], 10, #绿色, #蓝色, 1)
    计次循环尾 ()
```

Figure 18 Recognition button program

RESULTS

After the program runs, it will automatically pop up the control interface. Click the "Identify" button, and the system will automatically recognize and mark the foreign object. The adjustment of video playback speed may interfere with the accurate recognition of abnormal objects by the monitoring system. The recognition results of abnormal objects in the videos shown in Figures 19 and 20 were played at speeds of 25 and 40 frames per second, respectively. As shown in Figure 20, the identification box for marking foreign objects is not precise.

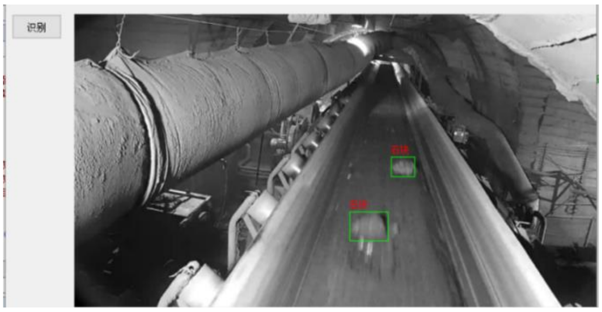


Figure 19 System foreign object recognition display (25fps)

Figure 20 System foreign object recognition display (40fps)

According to the calculation formula and result data in Section 4.2, the recognition performance of videos at different frame rates is compared, as shown in Table 1.

Table 1 Comparison of Recognition Results at Different Speeds

Video speed/fps	Total number of samples/number	accuracy/%	recall/%
25	400	88.5	91.6
40	400	85.4	88.7

When the frame rate of the video changes, it may cause the features of foreign objects to appear differently at different speeds, thereby affecting the accuracy of system recognition.

4.1.2 Detection of belt damage

Belt damage can have an impact on the normal operation and service life of the belt. The wear at the connection may lead to a decrease in the strength of the belt and make it prone to breakage, and it needs to be repaired or replaced in a timely manner. The attachment of foreign objects can increase friction and wear, which may lead to local wear and even breakage. The system identifies belt damage as shown in Figure 21.

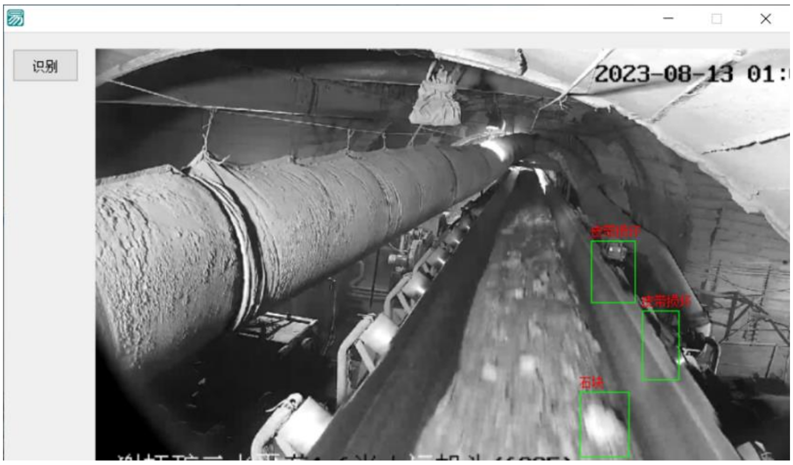


Figure 21 System identification of belt damage display

4.1.3 Model Validation and Optimization

Enter the model validation interface, place the model files trained 400 times and 10400 times in their corresponding positions, and then select images with foreign objects for recognition validation. The output results are shown in Figures 22 and 23, respectively. It can be seen that as the number of training iterations increases, the accuracy and generalization ability of the model in identifying foreign objects will improve.

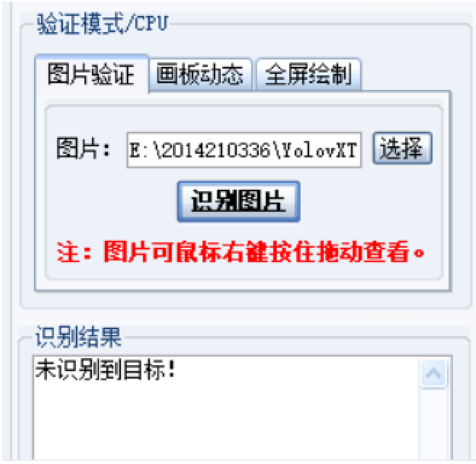


Figure 22 Model validation results (Iteration=400)

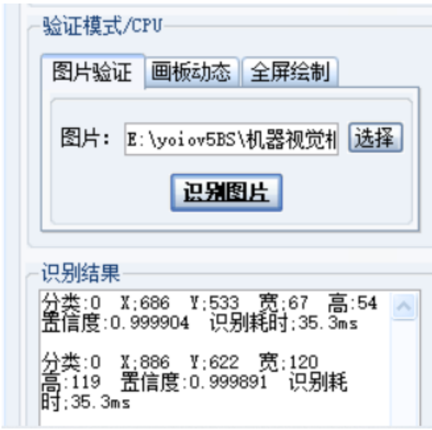


Figure 23 Model validation results (Iteration=10400)

4.2 System Performance Analysis

In order to further analyze the performance of the system, this article conducted statistical and analytical analysis on the experimental results:

Precision: The proportion of the test set that is correctly identified as positive samples by the model. The calculation formula is shown in (1):

$$\text{Precision}=\frac{TP}{TP+FP} \qquad (1)$$

Recall: The proportion of positive samples correctly identified as positive samples in the test set, calculated using the formula shown in (2):

$$\text{Recall}=\frac{TP}{TP+FN} \qquad (2)$$

Among them, positive samples refer to samples in which the target object exists in the image;

Negative sample: refers to samples in the image that do not contain the target object;

TP：TruePositive， The number of correctly identified positive samples;

FP：FalsePositive， The number of incorrectly identified negative samples;

FN：FalseNegative， The number of misidentified positive samples;

Table 2 Statistics and Analysis of Experimental Results

name	Quantity/piece	Correctly identify the number/piece	Number of misidentifications/piece
Positive sample	300	280	20
Negative sample	100	60	40
stone	500	485	15
Belt damage	20	19	1

The experimental results are shown in the statistics. By using the calculation formulas for accuracy and recall, it can be concluded that Precision is about 0.875 and Recall is about 0.93.

Overall, it can be concluded that high recall of true positive samples results in higher accuracy of the model's positive category prediction. Overall, this model can comprehensively cover the true positive categories (high recall rate) when identifying targets, and has high accuracy (high precision) when judging as positive categories. This result indicates that the model performs well in object detection tasks, and can comprehensively identify positive class samples while maintaining high accuracy.

DISCUSSION

This study proposes a method for identifying foreign objects in belt conveyors based on the law. Firstly, the model should be designed reasonably and a group pre training strategy should be implemented, while adopting various technical means to prevent model overfitting. The experimental process comprehensively covers data collection and annotation, data preprocessing, model training and optimization, model calling, and testing. The data preprocessing stage involves scaling, cropping, and normalization operations of belt conveyor images, aiming to meet the input requirements of the model. At the same time, data augmentation strategies have been implemented to enhance dataset diversity and model accuracy. Based on the annotation information of the dataset, generate corresponding training samples and labels, and reasonably divide the training set and the testing set. During the model training process, closely monitor performance indicators such as loss function and accuracy to ensure training effectiveness. Using pre trained weight files as initialization parameters, iterative training is carried out, and measures such as data augmentation, regularization, and early stopping are comprehensively applied to prevent model overfitting. Finally, in the programming environment of Easy Language, the trained YOLO model was successfully called to recognize the belt conveyor image as input and output accurate recognition results. According to actual needs, the recognition results can be further processed and analyzed. The experimental results comprehensively evaluated the system performance and effectiveness, and proposed relevant discussions and suggestions, providing useful references for subsequent improvement and optimization.

REFRENCES

[1] Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

[2] Zhu Jiarui, Wang Junjie Research and Implementation of Neural Network based Foreign Object Recognition Method [J]. Automation and Instrumentation, 2019, 40 (11): 63-67.

[3] Zeng Wenbing, Li Jun. Overview of deep learning based object detection algorithms [J/OL]. Automotive Engineer: 1-11[2024-01-04].

[4] Luo Yuan, Wang Boyu A review of research on object detection technology based on deep learning[J]. Semiconductor optoelectronics, 2020, 41(01): 1-10.

[5] Bai Ying Belt Conveyor Safety Detection System Based on Video Intelligent Recognition [J]. Modern Mining, 2023, 39 (6): 59-62.

[6] Wang Shanchao, Yu Feng, Sun Hongbao, etc Artificial Intelligence based Fault Diagnosis System for Conveyor [J]. Automation, 2023 (1): 131-133.

[7] Wang Yan Belt Conveyor Deviation Detection System Based on Sound Array and Image Processing [J]. Measurement and Control Technology, 2022 (006): 041.

[8] Zhao Y,Zhang L,Li M.Object recognition system design based on YOLOv5 and Easy Language[J].Journal of Computer Science and Technology,2020,48(6),123-129.

[9] Li Song, Dong Litao, Huang Lei, etc Transformation and Application of Capacity Enhancement for Belt Conveyor Process Funnel [J]. Chinese Science and Technology Journal Database (Full Text Edition) Engineering Technology, 2021 (3): 2.

[10] Zhu Jiarui, Wang Junjie Research and Implementation of Neural Network based Foreign Object Recognition Method [J]. Automation and Instrumentation, 2019, 40 (11): 63-67

[11] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[C]. Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9.Springer Berlin Heidelberg, 2006: 404-417.

[12] Song Jiajia, Zuo Chao Analysis of Causes and Maintenance Countermeasures of Coal Conveyor Belt Damage [J]. Enterprise Introduction, 2013 (16): 190.

[13] Wang X,Yang L,Zhang R,et al.Research on object recognition algorithm based on YOLOv5 and Easy Language[J].Journal of Electronic Technology Applications,2020,47(9), 98-103.

[14] Jingyuan Machine Vision based Coal Mine AI Tape Foreign Object Recognition Control System [J]. Energy and Energy Conservation, 2023 (4): 133-135.

[15] Yuan Haili, Xu Feng, Li Zhao, Wang Tao Belt Intelligent Inspection System Based on Visual Analysis [J]. Electronic Technology and Software Engineering, 2022,4 (23): 124-127.

[16] Lu Xueming, Yu Zaichuan, Xu Shengsheng Detection of foreign objects in coal mine transportation belts based on deep generative models [J]. Computer System Applications, 2022,6 (05): 358-363.

[17] Lin W B,Xu X S,Wang J G.Feature Selection for Image Recognition of Flaw Based on Genetic Algorithm[J].Journal of Measurement Science and Instrumentation,2016,27(10):25-28.

[18] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 2778-2788.