

Artificial intelligence garbage sorting vehicle based on deepstream

Chenxi Qi, Zhuoyi Wu, Luyao Wang, Honghua Feng, Hongyu Yang, Hongtao Liu*

Northwest Minzu University, Lanzhou, 730000, China

*Corresponding author: 625174277@qq.com

Abstract: In this paper, an intelligent garbage sorting vehicle based on the upper computer jetson nano and STM32 is designed. The upper computer jetson nano collects images of garbage objects through a usb camera, and the collected data is applied to the deep learning algorithm YOLO model to complete the garbage identification and classification function, and the radar is used to locate and avoid obstacles. Through the communication between the upper computer and the lower computer STM32, the trolley is driven to the specified position, and the garbage is sorted to the corresponding trash can by the mechanical arm. STM32 with the Internet of things function, can display the quantity and type of each garbage on the mobile app, so that the administrator can timely operation and maintenance. In this paper, the hardware and software design and joint debugging of the above system are completed from the reality.

Keywords: Garbage sorting; Image recognition; YOLO; The Internet of things

1. Introduction

With the rapid development of economy and the continuous improvement of people's living standards, the amount of garbage has surged in the great cycle of consumption and production. According to a research report by Lianhe Zaobao, the amount of garbage in the world will increase by 70 percent by 2050, from 2.01 billion tons in 2016 to 3.4 billion tons in 2050. The task of garbage sorting is very difficult, and it is urgent to research an effective garbage sorting method. Scholars at home and abroad have done a lot of analysis on garbage sorting, but most of the proposed schemes are the innovation of end recycling methods. Front-end collection depends on people's consciousness, and the efficiency of garbage sorting is very low. Studying an effective front-end collection method is of great significance for the current situation of garbage sorting [1].

For the front-end collection of garbage, the domestic and foreign markets began to appear intelligent garbage sorting, mostly based on infrared sensors, liquid crystal display screen design. It has the functions of automatic classification, alarm and monitoring, but it can not meet the needs of automatic identification and classification.

With the rapid development of artificial intelligence technology, machine vision technology is gradually mature, image classification method based on the deep learning gradually diversification[2], spam recognition for this project research provides a new direction, description is made valid by means of classification and recognition, choose different garbage as a sample under test, analyze the images of the garbage, and combined with image processing and pattern recognition technology, This paper proposes a garbage classification and identification method based on machine vision.

This paper proposes a garbage sorting system based on machine vision. The software system of garbage identification is designed, which can meet the requirements of garbage classification. The model training method is used to train the identifiable model, and the accuracy of the model is verified. The model is deployed on the Jetson nano development board, and the hardware structure is combined to complete the whole garbage sorting system.

2. Research method

2.1 Study the technical route

In the research of the project is based on the artificial intelligence Deepstream garbage sorting, the

garbage sorting car compared to the existing of garbage sorting equipment on the market, do the sorting and transport integration, the traditional garbage sorting is given priority to with manual sorting and sorting machine, the machine need sorting by the garbage truck transport to the designated venue in sorting, too much waste of time and resources. This garbage sorting vehicle uses YOLO algorithm to classify and process garbage, LiDAR for path planning, and robotic arm for garbage sorting, which greatly solves the problem of resource waste. At the same time, this garbage sorting vehicle can be connected to the Aricloud platform and mobile app to visualize data and facilitate data processing [3]. The project mainly includes the following parts: Jetson nano control module, manipulator control module, ESP8266 iot module, LiDAR, STM32 control module. At its heart is the Jetson nano control module. Jetson nano is a development board for deep learning or computer vision, which integrates 128-core Maxwell GPU, quad-core ARM A57 64-bitCPU and 4GLPDDR4 memory. Deepstream is a framework introduced by Nvidia specifically for AI processing of video streams, featuring GPU HW accelerated building blocks that bring deep neural networks and other complex processing tasks into the stream processing pipeline.

The combination of Jetson nano controller and Deepstream architecture and YOLO algorithm greatly improved the accuracy of garbage identification. In terms of path planning, the controller combined with LiDAR to construct a map for path planning, which was consistent with the path planning and use scheme of unmanned express vehicles, making the project more intelligent. Garbage sorting adopts the combination of upper computer programmed action group and PID algorithm to control the robot arm for sorting, which makes the sorting more accurate, convenient and efficient. In order to get more convenient data processing, the ESP8266 module is used to access Ali Cloud platform and mobile app through MQTT protocol, and the data is uploaded to the cloud platform and mobile app for data processing.

2.2 Convolutional neural network YOLO

The real-time object detection algorithm based on YOLO is a deep learning convolutional neural network model. In this model, object detection is regarded as the solution of a regression problem, and the methods such as BN (Batch Normalization), K-Means cluster evaluation anchor frame, backbone network and high-resolution classifier are used for network training, thus greatly accelerating the learning efficiency of the network [4].

The detection process of YOLO algorithm is shown in Figure 1. Firstly, the grid is segmented, and the original image is segmented into 7×7 grids. Then each grid generates two boundary prediction boxes, and the target is restricted and classified according to the credibility information. There are 98 prediction boxes in total. Secondly, the confidence information is calculated, and the confidence interval is calculated for each prediction block. Finally, the repeated blocks are deleted, and the prediction blocks are obtained by suppressing the non-maximum value.

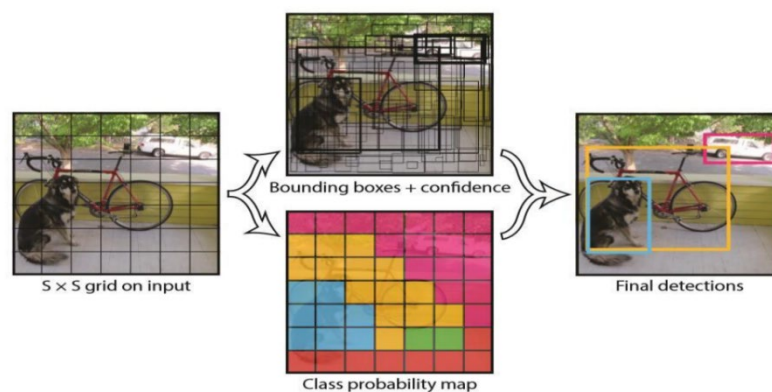


Figure 1: Schematic diagram of the detection process

2.3 Darknet-19

It is mainly composed of image acquisition, trunk feature extractor Darknet-19 and YOLO. Darknet-19 is mainly composed of a convolution module (DBL) and five residual layers. YOLO outputs three feature maps with different dimensions for multi-scale object detection. The first process is used for small-scale pedestrian target detection. On this basis, DBL is the basic component of YOLO, including

convolutional layer, BN layer and activation layer. Operation 1 is to perform five DBL operations on the feature map, and operation 2 is to perform DBL operations on the feature map and then upsampling operation [5].

Darknet-19 mainly uses 3*3 convolution. After using 2*2 maximum mapping layer, the dimension of feature map is reduced by 2 times and the channel of attribute map is doubled at the same time. Darknet-19 finally uses global avgpooling to make prediction, and uses 1*1 convolution method to carry out 3*3 convolution for feature map channels, so as to reduce the calculation amount and parameters of modeling. Batchnorm layer was used after each convolution in Darknet-19 to accelerate convergence and reduce pattern overfitting. In ImageNet, top-1 and top-5 reached 91.2% and 91.2%, respectively. After YOLO applied Darknet-19, the mAP value increased by 5%, and the computing power decreased by about 33%.

1) Extract the bounding box

To limit the center point of the border in the current cell, the offset is performed using the sigmoid function so that the resulting deviation is in the interval (0,1) (the scale of each cell is treated as 1). In summary, the actual size and position of the bounding box can be obtained by predicting the border of the four offsets [6], that is, calculating t_x, t_y, t_w, t_h .

$$\begin{cases} b_x = (\sigma(t_x) + c_x), b_y = (\sigma(t_y) + c_y) \\ b_w = p_w e^{t_w}, b_h = p_h e^{t_h} \end{cases} \quad (1)$$

Where (c_x, c_y) is the upper left coordinate of the cell, as shown in FIG. 5. During operation, the size of each cell is 1, so the upper left coordinate of the current cell is (1, 1). The sigmoid function limits its center to the interior of the current cell, thus avoiding excessive offset. While (p_w, p_h) is the width and length of the prior box. As mentioned before, their values are related to the size of the feature map. In the feature map, the length and width of each cell are both 1. Here, the size of the feature map is written as (W, H) , and then the edge block is calculated according to the size and position of the whole image: (the four numerical results are all between 0 and 1).

$$\begin{cases} b_x = (\sigma(t_x) + c_x)/W, b_y = (\sigma(t_y) + c_y)/H \\ b_w = p_w e^{t_w}/W, b_h = p_h e^{t_h}/H \end{cases} \quad (2)$$

Figure 2 shows the schematic diagram of bounding box calculation. If the above four values are multiplied with the width and length (pixel value) of the image, the position and size of the final bounding box can be obtained. That's the whole decoding process for YOLO. By predicting the restricted position of the border, we can better learn the stability. By integrating the clustering results with it, we can get a result that is suitable for it, and the result shows that the mAP of YOLO is improved by about 5% [7].

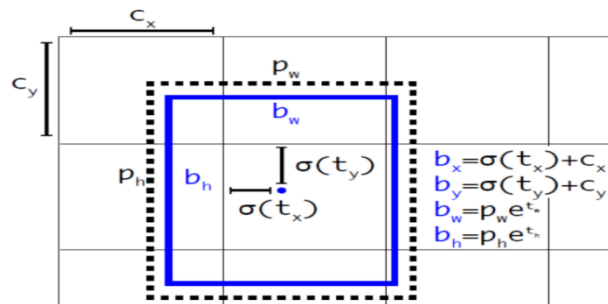


Figure 2: Schematic diagram of bounding box calculation

2.4 Fine grained feature

YOLO's image input resolution is 416*416. After 5 maxpooling (maximum pooling), 13*13 size images can be obtained, and then the feature map is used for preprocessing. A 13*13 feature map can be used to detect large targets, while smaller targets require more detailed feature maps. SSD uses graphs with multiple feature quantities to detect objects of different sizes individually. YOLO suggests using more detailed feature maps for detection. The detailed feature maps used in YOLO are 26*26, and the corresponding feature maps in Darknet-19 are 26*26*512. passthrough is combined with ResNet, taking the previous higher resolution feature map as an input and combining it with subsequent lower resolution feature maps. The dimension of the previous feature graph is twice that of the latter. The passthrough image extracts 2*2 local area from the previous layer and converts it to the channel size. The 26*26*512

feature graph is processed by the passthrough layer. The new feature map of $13 \times 13 \times 2048$ was obtained (the size of the feature map was reduced by four times, and the channel was increased by four times). The feature map of $13 \times 13 \times 3072$ was obtained by combining the feature map of $13 \times 13 \times 1024$, and the prediction was made based on the convolution of the feature map [8].

3. Garbage identification process

3.1 VoTT training set annotation

After verifying the successful configuration of the environment, you can create a dataset for local training. Data collection requires a large number of images to be collected, and the image resolution must be 224×224 . Otherwise, the image recognition will be inaccurate. Filming can be done manually using an Android phone or using python scripts. Once a third target is found, it is photographed and stored on an SD card. When conditions permit, a large number of pictures are taken to ensure the accuracy of the model [9].

Secondly, VoTT software is selected to label the training set. VoTT is a marking software for image objects launched by Microsoft, which is based on javascript and can be used on Windows, Linux, Mac and other different operating systems. Moreover, it also supports tagging from image to image and video to video, which is a good advantage to be able to count and visualize the labeled data. Most importantly, VoTT can be exported in a variety of formats, the export format is VOC.

After creating the project, click Source Connection: select the original data path of the photo. Click Target Connection: Select a path for storing the target data. Select the default values for other requirements. After the preceding operations are complete, you can start marking. Select the rectangular box for labeling, add data label press enter and a green label appears to complete labeling.

After the annotation is completed, you must click Save to ensure that the output format is the xml format required by yolo, and select export to the VOC format required to complete the annotation of the whole dataset.

3.2 Garbage image classification and recognition function

Before training, pre-training is carried out on ImageNet. The pre-trained classification model adopts the first 20 convolutional layers in Figure 3, and then adds an average-pool layer and a fully connected layer. After pre-training, four randomly initialized convolutional layers and two fully connected layers are added to the 20 convolutional layers obtained by pre-training. Since the detection task generally requires higher resolution images, the input of the network is increased from 224×224 to 448×448 .

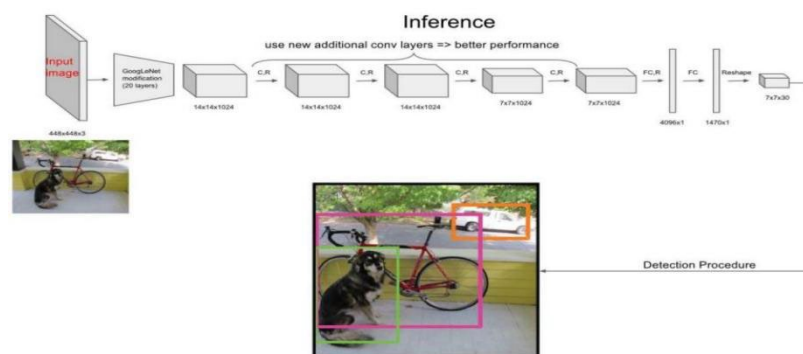


Figure 3: YOLO Web Design

3.3 Model prediction

First, I will introduce the non maximum suppression (NMS) algorithm. This algorithm is not only for Yolo algorithm, but is used in all detection algorithms. NMS algorithm mainly solves the problem that a target is detected multiple times. As can be seen from face detection, face is detected multiple times, but we actually hope to output only one of the best prediction boxes in the end. NMS algorithm can be used to achieve such effect: Firstly, the box with the highest confidence is found from all the detection boxes, and then its IOU with the remaining boxes is calculated one by one. If its value is greater than a certain

threshold (the coincidence degree is too high), the box will be removed. The process is then repeated for the remaining detection boxes until all detection boxes are processed as shown in Figure 4. NMS algorithm is also needed in the Yolo prediction process [10].



Figure 4: Battery test chart and food bag test chart

4. Device integration

4.1 Hardware structure design

Familiar with the driving principle of STM32F1 series microcontroller. The functions used in this project include I/O port reuse, IIC communication, PWM wave output, serial communication, timer interrupt, external interrupt, timer and other functions. Familiar with Jetson nano and Linux operating system, and familiar with the use of its modules, can skillfully use it to control other modules, familiar with its driving principle. And understand some image processing algorithms about object detection, can carry out simple filtering processing, understand the principle of network communication. On this basis, the software and hardware of this project are designed, and the functions used by each module are matched one by one, and the implementation steps and schemes are made.

4.2 Software programming

The programming language used in this project is C language, and the compiler used is Keil5. Keil compiler perfectly supports the use of C language to STM32 series chip driver, the compiler function is very powerful, support online simulation function, you can see the value of each register through the observation window, you can more intuitively see the program design is the error and timely correction, for the debugging of the program is very helpful. When writing the program, modular programming is used to encapsulate the initialization and driving process of different modules in different.c files. The biggest advantage of modular programming is that the program is more readable, the embodiment of programming ideas is more clear, and it is also very convenient for the debugging and modification of each module of the program. To build hardware and conduct evaluation and analysis of various environments, to construct software and test a variety of environments.

In order to achieve the accuracy of garbage identification, this project uses a high-definition camera and optimizes the image algorithm on this basis, so that the real-time detection image can reach about 30 frames and the target detection can be carried out accurately. The connected Jetson nano was programmed to achieve the purpose of accurate recognition, and the processing results were fed back to STM32. Many modules are used in this project, and different communication protocols are used in the use of different modules. Therefore, in the circuit design, different modules should be connected to appropriate IO ports according to the hardware resources of different IO ports, so as to realize the reasonable use of hardware resources.

The design of the reasonableness and stability of the circuit of the control system of the hardware structure of the whole system is a very key step. We can use simulation software to realize the design, so that there will be no repeated modification. After confirming that the desired effect is achieved, we can start to build the hardware as a whole. This allows a lot of time to debug the software.

The possible problems in different simulation situations are studied, and the functions of the garbage sorting vehicle are constantly optimized and improved, so as to improve the adaptability of the garbage sorting vehicle in different environments.

5. Conclusion

In this paper, jetson nano is used as the upper computer, STM32 development board is used as the lower computer, combined with YOLO algorithm and other technologies to design and develop an intelligent garbage sorting vehicle. The system provides ultra-accurate image recognition technology, so as to realize automatic garbage identification and sorting. The system is also equipped with a variety of embedded modules to operate and maintain the control system. The research content of this paper is summarized as follows:

(1) Firstly, this paper introduces the development status of the application of deep learning and image recognition technologies in intelligent garbage classification system, discusses the current situation of garbage classification in the world and relevant policies and systems formulated by various governments, and studies the current intelligent garbage classification system.

(2) Analyze and demonstrate the overall design scheme of the system, and complete the hardware selection and hardware circuit design.

(3) Analyze the core technology of this paper -- YOLO model, introduce the structure and defects of this model, and then improve and optimize the structure, so as to obtain the YOLO model with higher recognition accuracy.

(4) For the built hardware platform, darknet is used for model training; After completing the coding of some STM32 modules, the upper computer jetson communicates with the lower computer STM32 through serial port, so that STM32 can receive the data after image processing. Finally, the ESP8266 is used to communicate with the lower computer STM32 through serial port to display corresponding data and realize the operation and maintenance management of the system.

(5) Firstly, the accuracy of YOLO model is verified, and the recognition accuracy is above 95%. Then the function of STM32 of the lower machine is debugged. The garbage sorting function of the robot arm of the lower machine is verified and can be controlled stably. Finally, ESP8266 uploads the data to the mobile app through MQTT protocol to facilitate the administrator's operation, maintenance and management.

Acknowledgement

The authors gratefully acknowledge the financial support from 2022 Northwest University for Nationalities College Students Innovation and Entrepreneurship Training Program Grant (Project No.202210742013) .

References

- [1] Rabano S L, Cabatuan M K, Sybingco E, et al. Common Garbage Classification Using MobileNet [C].10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE 2018: 1-4.
- [2] Chen Shasha. Research on Recyclable Garbage classification Algorithm based on deep learning [D]. Tutor: Du Qingdong. Shenyang Normal University, 2021.
- [3] Changhu Zhou, Zhenyu Chen, Xiaotian Lv, Derun Gao, Mengxue Zhao. Design of Intelligent Sorting Trash Dustbin Based on STM32 [J]. E3S Web of Conferences, 2020, 198(198):30-34.
- [4] Yang M, Thung G. Classification of trash for recyclability status [J]. CS229 Project Report, 2016, 2016.
- [5] Awe O, Mengistu R, Sreedhar V Smart trash net: Waste localization and classification [J]. arXiv preprint, 2017.
- [6] Ruiz MSanchez A, Velez J F, et al. Automatic image-based waste classification[C]//International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, Cham, 2019: 422-431.
- [7] Aral R A, Keskin SR, Kaya M, et al. Classification of trashnet dataset based on deep learning models[C]//2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018: 2058-2062.
- [8] Connor Shorten, Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning [J]. Journal of Big Data, 2019, 6(1):13-17.
- [9] Minaee Shervin, Boykov Yuri Y, Porikli Fatih, Plaza Antonio J, Kehtarnavaz Nasser, Terzopoulos Demetri. Image Segmentation Using Deep Learning: A Survey. [J]. IEEE transactions on pattern

analysis and machine intelligence,2021, PP:23-27.

[10] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. *Neural Machine Translation by Jointly Learning to Align and Translate [J]*. CoRR, 2014, abs/1409.0473(abs/1409.0473):20-24.