# YOLOv5-based Student Counting Software Design

**Jie Rao[1,*], Mingju Chen[1, 2]**
*[1]College of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin, Sichuan, China*
*[2]Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin, Sichuan, China*
**Corresponding Author*

**Abstract: To address the issues of students' campus safety as well as campus management, this paper combines computer vision technology and deep learning algorithms to design a counting software for observing the number of students in a classroom, which can provide accurate headcount and data analysis by monitoring and counting the classroom crowd in real time in order to solve the problem of classroom crowd counting. The classroom crowd counting software based on YOLOv5 has a wide application potential in the field of education. It provides real-time and accurate headcount statistics for classroom management and supports them in making decisions on staff scheduling and resource management. In this paper, we adopt YOLOv5 algorithm as the main target detection framework, which is capable of fast and accurate target detection and localization. Then, this paper designs a crowd counting software based on Qt Designer, which can monitor the number of people in the classroom in real time and perform accurate headcount. In addition, we added data visualization and analysis functions to the software for more in-depth analysis of the headcount results. Finally, experiments on the publicly available benchmark dataset CUHK Occlusion show that the algorithms in this paper exhibit significant advantages in terms of accuracy and real-time performance.**

**Keywords: Classroom Student Counting; Yolov5 Algorithm; Qt Designer; Target Detection**

## 1. Introduction

Students' campus safety as well as campus management are always hot topics. In the field of education, accurately monitoring and counting the number of people in the classroom is crucial for improving teaching efficiency, optimizing resource allocation, and ensuring students' safety. To meet the needs of classroom crowd counting, this paper designs a classroom crowd counting software based on the YOLOv5 algorithm to achieve the goal of real-time monitoring and accurate counting of the number of people in the classroom.

By adopting the YOLOv5 algorithm as the target detection framework, this paper solves the problem of classroom crowd counting by combining it with computer vision technology while making full use of its fast and accurate target detection and localization capabilities. The crowd counting software designed in this paper is not only able to monitor the number of people in the classroom in real time, but also provides data visualization and analysis functions, which enable classroom managers to have a deeper understanding of the statistical results and make corresponding scheduling and management decisions.

In this paper, the design and implementation of the classroom crowd counting software based on YOLOv5 is described in detail, as well as the evaluation and analysis of its performance and practicality. The results show that the proposed method has significant advantages in terms of accuracy and real-time performance, providing a practical and efficient crowd counting solution for the education field.

## 2. Related Technologies

### 2.1 Introduction to Target Detection Algorithms

Target detection algorithms are one of the important techniques in the field of computer vision to automatically identify specific target objects from images or videos. These algorithms use deep neural networks to analyze the input image data to extract key features. Through the design and training of the network structure, they are able to realize the target detection task and decompose it into a series of subtasks. The three basic steps of a target detection algorithm are: image classification, segmentation and target localization [1].

## 2.2 Introduction to the YOLOv5 Algorithm

### 2.2.1 YOLOv5 network architecture

The core components of the YOLOv5 algorithm include Backbone network, Neck network and Head network. Their respective functions are shown in Table 1.

**Table 1. Network Menu**

| Network | Functionality |
|---|---|
| Backbone | Extraction of raw features |
| Neck | Further optimized on Backbone |
| Head | Extracting target information from optimized features |

In the Backbone network, the presence of Focus structure and CSP structure can be seen.The Focus structure serves to channel-split the input feature maps and recombine these channels so that the network focuses more on smaller regions to improve the detection precision and accuracy of small targets [2].The CSP structure essentially divides the whole network into two parts and uses cross connectivity to divide the feature map into two halves. The feature maps of these two parts are then cascaded.The structure of YOLOv5 is schematically shown in Figure 1.

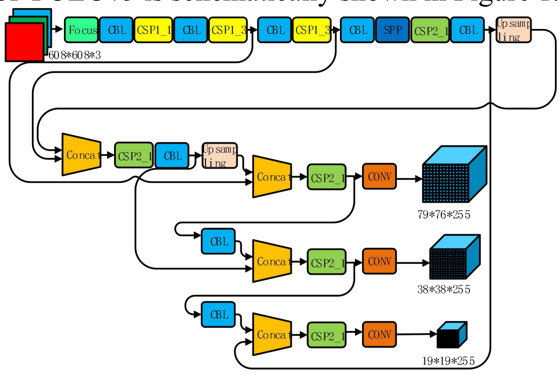


**Figure 1. YOLOv5 Structure Schematic**

The above figure is quoted and reformatted from jiangdabai. In YOLOv5, the CSP structure is used in a way that can better capture the rich information in the image, thus further improving the accuracy and efficiency of target detection.

### 2.2.2 Core network

In this paper, the backbone feature extraction network used in algorithm YOLOv5 is CSPDarknet, which uses the residual network Residual, the residual convolution in CSPDarknet can be divided into two parts, the backbone part is a one time 1×1 convolution and a one time 3×3 convolution; the residual edge part is not processed in any way, and combines the inputs of the backbone with the outputs directly [3].

It effectively solves the problem of gradient vanishing by means of jump connections and residual blocks, and improves the depth and feature expression of the network, thus realizing better performance and effect of target detection. The design of residual network in CSPDarknet effectively alleviates the training difficulty of the deep network while extracting richer features, and retains the original input information by means of the structure of jump connections and residual blocks, which It makes the network better learn the fine-grained feature information and improves the accuracy and efficiency of target detection.

### 2.2.3 Focus structure

YOLOv5 introduces the Focus structure, a special design that plays a key role in processing the input image for feature extraction. Specifically, it operates by cropping and splitting the input feature map into four sub-maps and interleaving and stacking these four sub-maps [4].The Focus structure effectively reduces the complexity and the number of parameters of the network. The Focus structure plays an important role in the YOLOv5 algorithm,this design not only improves the efficiency and accuracy of the model, but also enhances the detection of small targets.

Its specific operation is to get a value every other pixel in a picture, at this time, four independent feature layers are obtained, and then the four independent feature layers are stacked, at this time, the width and height information is centralized to the channel information, and the input channel is expanded four times. The Focus structure divides the input feature map into several smaller regions, and then performs convolution operations on each small region to obtain the corresponding feature representation. Subsequently, these feature representations will be concatenated or fused together to form the final output feature map. This design can effectively reduce the computational complexity of the network and improve the detection accuracy of the network for small targets.

2.2.4 Mosaic data enhancement

In YOLOv5, Mosaic data augmentation is an effective data augmentation technique used to increase the diversity and quantity of training data to improve the robustness and generalization of the target detection model. Mosaic data augmentation creates a new synthetic image by combining multiple different image segments so that the model learns to deal with different scenarios and transformations of the target during training.

Mosaic data augmentation randomly selects four different images and crops, reorganizes, and stitches them into a new training sample. By introducing Mosaic data augmentation, the model is exposed to more diverse data during the training process, which improves its ability to adapt to different scenarios, thus increasing the accuracy and robustness of detection. Here, four classroom student images are selected as shown in Figure 2.

**Figure 2. Pictures of Students in Classrooms**

After cropping, reorganizing, and splicing through Mosaic data enhancement, it becomes a training sample of a picture, and the target detection results obtained through training enhancement are shown in Figure 3.

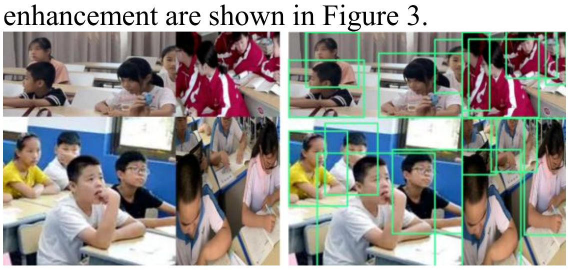

**Figure 3. Training Results Chart**

Enhancement with Mosaic data improves the performance and generalization of the model in target detection tasks.

2.2.5 Multiple positive sample matching

Multi-positive sample matching is a key technique in target detection in YOLOv5, which is used to effectively deal with the problem of matching multiple target objects existing in the same lattice [4]. Usually, there may exist multiple target objects in a lattice, and traditional target detection algorithms tend to match only one of the targets in this case, resulting in the problem of missed or false detection.

Therefore, in order to improve the accuracy and effectiveness of target detection, YOLOv5 introduces the concept of multiple positive samples matching, which allows multiple target objects in a lattice to be used as positive samples to participate in the loss calculation and gradient update [5]. In this way, during the training process, the model can better learn the relationships and features between different targets, improving the recognition and detection of multi-target scenes.

In order to realize multiple positive samples matching, the following four key points are to be included in the implementation approach, which are: matching strategy, matching degree, localization accuracy and gradient update.

### 2.3 Introduction to Counting Software

In the design of the technical software, the use of Qt Designer is taken to design our technical software. The principle is to design the interface of the software in Qt Designer, by converting it to python code, and then write the logic code of each board, so as to realize the perfect combination of interface and function.

## 3. Methodology and Results

### 3.1 Model Training

The dataset used for the training of this model is from CUHK Occlusion Dataset, which is used for the study of activity analysis and congestion scenarios, and contains 1063 images of occluded pedestrians, of which 955 images are used as the training set and 108 images are used as the test set in this paper. After the training, the labeled samples as well as the training results can be obtained, the labeled samples are shown in Figure 4 and the training results are shown in Figure 5.

In target detection model evaluation, Recall and Precision are commonly used evaluation metrics.

Recall represents the coverage of the model, while precision represents the accuracy of the model, and they take values ranging from 0 to 1, with values close to 1 indicating better performance and close to 0 indicating worse performance [6]. In order to fully assess the performance of the model, it is usually evaluated using mean average precision (mAP). mAP calculates the corresponding precision and recall by setting different confidence thresholds and plots a curve based on the p-value and r-value under different thresholds. The area under the curve is the average precision (AP) of each category, and the AP values of all categories are averaged to obtain the mAP value of the model, which is used to comprehensively evaluate the performance of the target detection model [7]. As shown in the experimental results in Figure 3, after setting different confidence thresholds, different results can be obtained, and after debugging to the best training results of this design, then proceed to the subsequent steps [8].

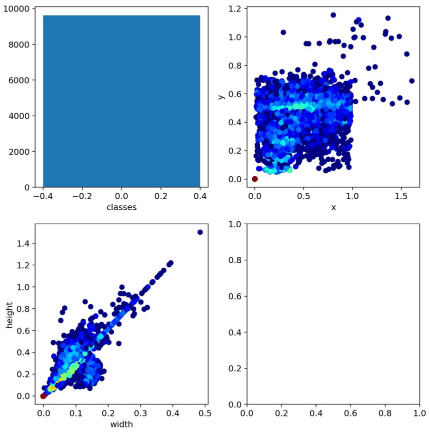


**Figure 4. Sample Label**

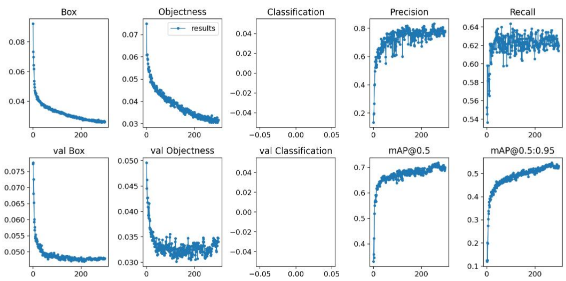


**Figure 5. Training Results**

### 3.2 Implementation of Counting Software

The counting software in this study is realized through Qt Designer, the basic idea is: design the interface of the counting software in Qt Designer, including the specific keys to be used and the display interface, after designing the software, save it as a .ui file, because this model is realized through python code, so by executing the code (1) at the command line to complete the .ui file to .py

pyuic5 input.ui -o output.py (1)

Where input.ui is the designed interface file and output.py is the output python file, after that, you only need to write the connection between the signal and each slot in the generated. py file [9], associate the events or actions on the interface with the back-end logic, and write the corresponding code to handle the interface operation and the connection logic of the training model. The perfect combination of interface and functionality can be realized.

In this article, the counting software interface mainly includes functions such as image detection, video detection, and camera detection. It can also indicate the time required for detection, the number and density of detection targets, and other information related to various detections. The specific implementation will be shown in the following text.

### 3.3 Results

Classroom crowd detection in the actual application, should not be limited to the detection of pictures, considering its scope of application, but also to meet the detection of video and real-time detection through the camera, in the detection of pictures, we use different classroom pictures to compare, in order to test the model of the good and bad, through the detection of pictures of classroom crowds as shown in Figure 6.

**Figure 6. Picture Detection Results**

Then the detection result of the video, the detection result of the video we will go through the frame analysis, and its detection result is shown in Figure 7.

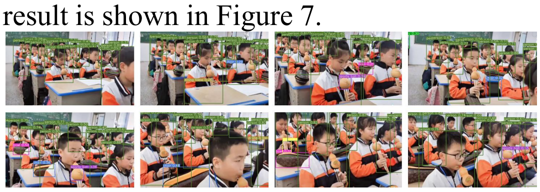

**Figure 7. Video Detection Results**

Given that modern classrooms are basically equipped with monitoring, in order to make it easier for administrators to manage students as well as to ensure student safety, this study also includes a camera module, which means that real-time monitoring can be carried out according to the camera, which can be satisfied with camera detection through previous training, and the results of the camera detection are shown in Figure 8.

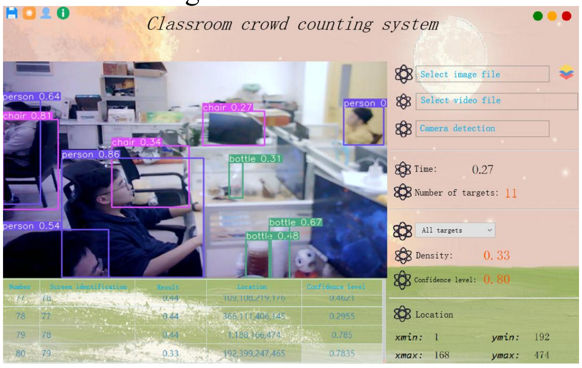


**Figure 8. Camera Inspection Results**

### 3.4 Analysis of Test Results

By further analyzing the experimental results, we can see that the YOLOv5 model has shown considerable accuracy in classroom crowd detection. However, we have also identified some challenges and areas for improvement. Especially in situations where there is extensive occlusion or poor image quality, the model may produce anomaly detection, which may affect the stability and reliability of the system.

We can adopt various strategies to improve the accuracy and robustness of detection in response to these issues. Firstly, we can consider optimizing the performance of the YOLOv5 model in complex scenarios by adjusting its parameters. For example, adjusting the threshold for object detection, changing the input resolution of the model, or adjusting the network structure may all help improve the performance of the model [10]. By continuously optimizing model parameters, combining with other algorithms, and utilizing various technical means, the detection accuracy of this model can be effectively optimized.

For problems where other labels appear in the detection, they can be corrected by changing the number of labels and categories in the code, which is not described here [11].

## 4. Summarize

In this paper, we design and implement a classroom crowd detection and counting software by using YOLOv5 algorithm and Qt Designer, which realizes the functions of picture detection, video detection and camera detection. Through the detailed introduction of the overall architecture, algorithm selection, interface design and functional modules of the software, we have made a comprehensive summary and analysis of the design and implementation of the software.

In this paper, YOLOv5 algorithm is adopted as the core algorithm of crowd detection. Based on the YOLOv5 algorithm, the software can quickly and accurately detect and count the crowd in the classroom. At the same time, we use Qt Designer for interface design to provide users with an intuitive and friendly interactive interface.

The software designed in this paper realizes the functions of picture detection, video detection and camera detection, which can meet the needs of different scenarios. Users can import a single picture for crowd detection and counting, or select a video file for real-time monitoring, or even connect a camera to realize real-time crowd detection in the classroom. Finally, we conducted a comprehensive performance evaluation of the software. By testing the software in several real classroom scenarios, the software shows high detection accuracy and real-time performance.

Classroom crowd detection based on the YOLOv5 algorithm has a broad application prospect in educational environments and is important for improving classroom safety management and crowd control. Future work can further optimize the algorithm and improve the performance and stability of the software to meet more needs and scenarios of applications.

**Acknowledgments**

This research was funded by the Natural Science Foundation of Sichuan (grant number 2023NSFSC1987)

**Reference**

[1] Sun J, Wang Z. Vehicle and Pedestrian Detection Algorithm Based on Improved YOLOv5. IAENG International Journal of Computer Science, 2023, 50 (4):
[2] Zhou J, Su T, Li K, et al. Small Target-YOLOv5: Enhancing the Algorithm for Small Object Detection in Drone Aerial Imagery Based on YOLOv5. Sensors, 2023, 24 (1):
[3] Shi L, Lu B, Qiu Y, et al. Hepatic mosaic enhancement pattern correlates with increased inflammatory activity and adverse therapeutic outcomes in patients

 

with Crohn's disease. Abdominal Radiology, 2021, 46 (7): 1-10.

[4] Mingju C, Zhongxiao L, Zhengxu D, et al. HDS-YOLOv5: An improved safety harness hook detection algorithm based on YOLOv5s. Mathematical biosciences and engineering: MBE, 2023, 20 (8): 15476-15495.

[5] Mingju Chen, Hongyang Li, Hongming Peng, Xingzhong Xiong, Ning Long. HPCDNet: Hybrid position coding and dual-frquency domain transform network for low-light image enhancement. Mathematical Biosciences and Engineering, 2024, 21(2): 1917-1937. doi: 10.3934/mbe.2024085

[6] Mingju Chen, Fuhong Qiu, Xingzhong Xiong, Zhengwei Chang, Yang Wei, et al. BILSTM-SimAM: An improved algorithm for short-term electric load forecasting based on multi-feature. Mathematical Biosciences and Engineering, 2024, 21(2): 2323-2343. doi: 10.3934/mbe.2024102

[7] Song F, Li P. YOLOv5-MS: Real-Time Multi-Surveillance Pedestrian Target Detection Model for Smart Cities. Biomimetics, 2023, 8 (6):

[8] Chen Y, Wang C, Zhang C. A Robust Lightweight Network for Pedestrian Detection Based on YOLOv5-x. Applied Sciences, 2023, 13 (18):

[9] Kanghui Z, Biaoxiong X, Xingang M, et al. LPO-YOLOv5s: A Lightweight Pouring Robot Object Detection Algorithm. Sensors (Basel, Switzerland), 2023, 23 (14):

[10] Şirin M G, Gültekin I. A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models. Journal of real-time image processing, 2023, 20 (1): 5-5.

[11] Yitong M. A pedestrian detection algorithm for low light and dense crowd Based on improved YOLO algorithm. MATEC Web of Conferences, 2022, 355.