ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

Design of a Sichuan Cuisine Dish Image Recognition and Nutritional Information Assessment System Based on YOLO v5s

Chen Xieyu[1,2,a,*], Tang Na[1,2,b]

[1]*School of Business, Geely University of China, Chengdu, China*
[2]*Key Laboratory of Sichuan Cuisine Artificial Intelligence, Chengdu, China*
[a]*chenxieyu@guc.edu.cn*, [b]*tangna@guc.edu.cn*
[*]*Corresponding author*

Abstract: *With the development of deep learning technology and increasing awareness of dietary structures, food recognition has become a popular research field. This article presents an integrated system that combines Sichuan cuisine dish recognition with nutritional assessment, based on the lightweight YOLO v5s model. The system consists of three main modules: data preprocessing module, model training and evaluation module, and nutrition information display and reminder module. In this study, a total of 1,602 images of Sichuan dishes were utilized as the original dataset, which were annotated to generate six categories: Yu Xiang Rou Si, Shui Zhu Rou Pian, Mapo Tofu, Qing Jiao Chao Rou, Gong Bao Ji Ding, and Kuo Shui Huang Gua. The results of the model's performance indicate that the mean Average Precision (mAP@0.5) value of the model is 0.97, with the prediction, recall, and F1-score for each class exceeding 0.9. The loss function exhibited stable performance, indicating that the model operates effectively. The article concludes by showcasing some recognition results of the dishes along with their nutritional composition information.*

Keywords: *YOLO v5s, Sichuan Cuisine Dish, Nutritional Assessment, Image Recognition*

1. Introduction

Currently, only 15% of the population in China is in a healthy state, while about 70% remains in a sub-healthy condition, often accompanied by one or more unhealthy dietary habits, such as irregular eating patterns and imbalanced nutritional intake. The detection rate of sub-health conditions among residents with such eating habits is significantly higher than that of those with a normal diet[1]. Sichuan cuisine is generally high in oil and salt. According to the World Health Organization (WHO), a high-sodium and high-fat diet increases the risk of cardiovascular diseases, stomach cancer, obesity, and kidney diseases. As public concern for dietary structure increases and target detection technologies mature, the application of food nutritional component recognition is gradually rising. The YOLO[2] (You Only Look Once) series of algorithms holds a significant position in the field of target detection, prized for its fast recognition speed, high accuracy, and strong real-time capabilities, making it popular among researchers and practitioners. With the continuous development of deep learning techniques, the YOLO series has now iterated up to YOLO v10. Compared to other YOLO algorithms, YOLO v5s is a lightweight model developed using PyTorch, with significantly reduced parameter and computation requirements, making it suitable for resource-constrained devices and facilitating secondary development or model fine-tuning for developers. Most importantly, it achieves a good balance between speed and accuracy, particularly excelling in real-time application scenarios where it can provide rapid inference speeds with lower hardware requirements[3]. YOLO v5s has been widely applied in various fields, including industry and agriculture. For example, Pan Zhang and Daoliang Li applied YOLO v5s to detect the survival rate of rapeseed seedlings at multiple growth stages in a plant factory, constructing a target detection model for the rapeseed seedling dataset and achieving strong model performance[4]. Researchers like Yin H and others utilized YOLO v5s in conjunction with attention mechanisms and improved upsampling algorithms to address the challenge of detecting light smoke in the early stages of a fire, thus aiding early fire warning systems. To solve the issue of untimely concrete crack inspections, Yu and others employed the YOLO v5s model to train and test 3,500 manually annotated crack images, enhancing the model's crack detection capabilities[4]. A review of the literature reveals that most studies focusing on dish image detection predominantly utilize YOLO v3, with few researchers employing the

ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

lightweight YOLO v5s model. Furthermore, there is a scarcity of studies that combine dish recognition with nutritional component assessment. This article designs an integrated system for Sichuan cuisine dish recognition and nutritional assessment based on YOLO v5s, aiming to simplify the dish recognition process, improve detection efficiency, and provide an easy access point for individuals seeking nutritional information about dishes. Additionally, it offers a reference for future researchers regarding model training and testing."

2. System Design

The deep learning neural network algorithm system based on YOLO v5s is designed with an integrated lightweight framework model. Users can upload images, and the system uses the YOLO v5s lightweight model to perform visual detection, classification, and nutritional information prompts for the target images. The system primarily consists of a data preprocessing module, a model training and evaluation module, and a nutritional information display and reminder module. The main workflow is illustrated in Figure 1.

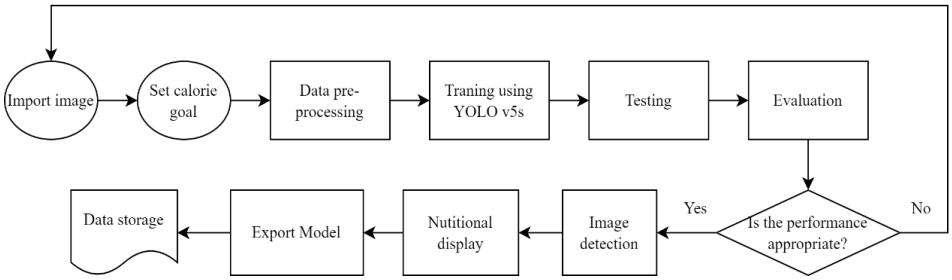

Figure 1: System working flow

2.1. Data pre-processing

Data preprocessing involves tasks including annotating images, dividing the image datasets, resizing the images, and applying image augmentation techniques.

2.1.1. Data collection

This study collected a total of 1,602 images of Sichuan dishes, with data primarily sourced from Baidu Images and the Roboflow. The nutritional dataset is sourced from the FatSecret website (https://www.fatsecret.com/) and includes information on energy, fat, carbohydrates, and protein. This data is primarily used to evaluate the nutritional value of dishes in each class.The selected dish datasets consist of home-style Sichuan dishes, including six classes: Yu Xiang Rou Si (Fish Fragrant Shredded Pork), Shui Zhu Rou Pian (Sichuan Boiled Pork), Mapo Tofu, Qing Jiao Chao Rou (Stir-fried Pork with Green Peppers), Gong Bao Ji Ding (Kung Pao Chicken), and Kuo Shui Huang Gua (Cucumber in Sauce). Each dish class contains 267 images.

2.1.2. Data annotation

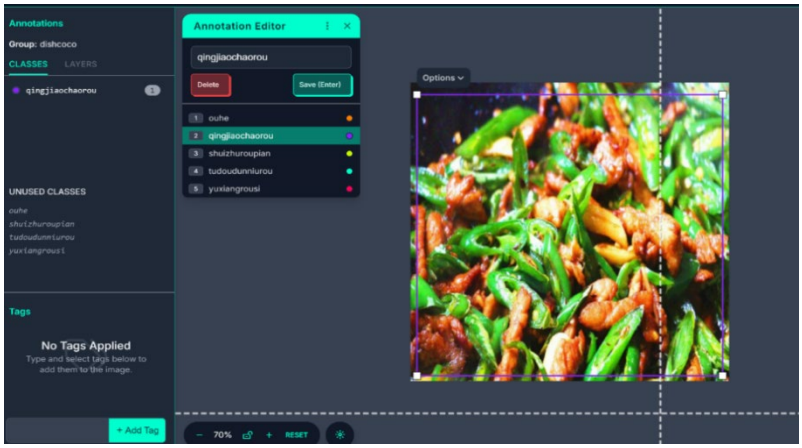

Figure 2: Data annotation

ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

Image annotation involves the process of labeling images to provide detailed descriptions that computers can understand. During annotation, different elements within an image, such as objects or specific areas, are identified and tagged with relevant labels. The article uses Roboflow as the image annotation tool, as shown in Figure 2. This tool simplifies the process of quickly annotating the locations and categories of target objects within images. The data annotation process using Roboflow primarily involves four steps. First, the dataset is imported into the project. Next, the Bounding Box Tool or Polygon Tool is used to locate the target objects. Then, the objects within the bounding boxes are annotated. Finally, while annotating, a text file with the .txt extension is generated, which contains the position and label information for the annotations.

2.1.3. Split datasets

To ensure the predicting accuracy, the original data must be divided into training, validation, and testing datasets[5]. This article uses a 70:20:10 ratio to split the images. In each class of 267 images, there are187 images for training data, 53 images for validation data, and 27 images for testing data.

2.1.4. Resize and augment the image

Image augmentation helps improve the robustness and generalization ability of models. Common image augmentation techniques include rotation, flipping, cropping, brightness adjustment, color jittering, and elastic transformations[6]. In this study, each class will include flipping, resulting in a new dataset that has increased the capacity of the original dataset; therefore, the new dataset will consist of a total of 2348 images.

2.2. Detection Model Based on YOLOv5s Algorithm

2.2.1. Model architecture

The structure of the YOLOv5s model is illustrated in Figure 3 and is composed of three main parts: the Backbone, the Neck, and the Output. The Backbone includes the Focus layer, CBS modules, CSP modules, and SPP modules, which is responsible for feature extraction from the input images, leveraging modules connections to improve gradient flow and reduce computational complexity. The Neck section utilizes a Path Aggregation Network (PANet) which enhances feature fusion by combining features from different layers of the backbone. The detection Head of YOLO v5s predicts bounding boxes and class probabilities. It employs multiple detection layers, typically three different scales, to detect small, medium, and large objects in the images.

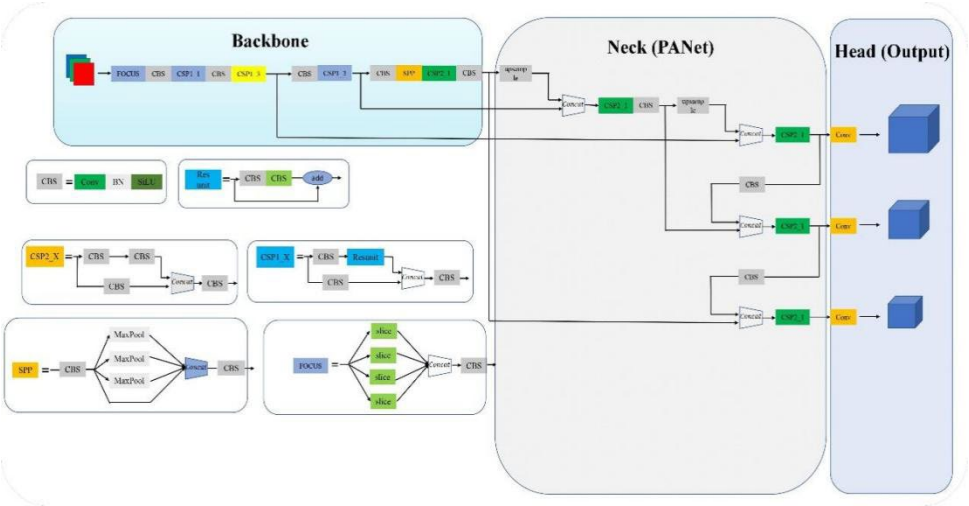

Figure 3: YOLO v5s architecture

2.2.2. Loss function

In YOLO v5s, the loss function can be subdivided into three main types: Box Loss, which measures the discrepancy in bounding box predictions; Object Loss, which assesses the confidence of object presence; and Class Loss, which evaluates the accuracy of the predicted classes[7].

(1) Box Loss

In general, YOLO v5s uses CIoU (Complete IoU), DIoU (Distance IoU), or other similar methods to

ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

calculate box loss. CIoU_loss algorithm enhances the ability to accurately and reliably detect food items by considering various geometric factors, which is critical in applications like automated food tracking, inventory management, and nutritional analysis[8].

(2) Class Loss

Class loss measures how accurately the model classifies the detected objects into their respective classes. This helps to ensure that once an object is detected, it is identified correctly. Class loss is generally calculated using categorical cross-entropy for each detected object[9].

(3) Object Loss

Object loss measures how well the model predicts the presence of an object in a given bounding box. This is particularly important because it helps the model understand not just where objects are but also whether the predictions contain actual objects. Object loss in YOLO v5s is usually calculated using binary cross-entropy (BCE) for each anchor box in the grid cell. The loss function measures the difference between the predicted probability of an object and the ground truth[9] .

2.2.3. Evaluation criteria

To measure the accuracy of the algorithm presented in this article, we primarily calculate precision, recall, andF1-score. The formula shown as below:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{3}$$

Where TP (True Positive) refers to the correctly identified positive samples; FP (False Positive) refers to the incorrectly identified positive samples; FN (False Negative) refers to the incorrectly identified negative samples. To further evaluate the performance of the improved model, we calculate the area under the precision curve and the coordinate axes (average precision, AP), as well as the area under the recall curve and the coordinate axes. By averaging these values, we can obtain the mean average precision (mAP) and mean recall (mRC), with both metrics ranging from 0 to 1.

2.3. Nutritional Assessment Module

The nutrition assessment module primarily focuses on the subsequent processing of the identification results from YOLO v5s. The nutritional analysis includes the content of protein, fat, carbohydrates, and the analysis of total calorie count. It includes the following three sub-modules, as shown in Figure 4.

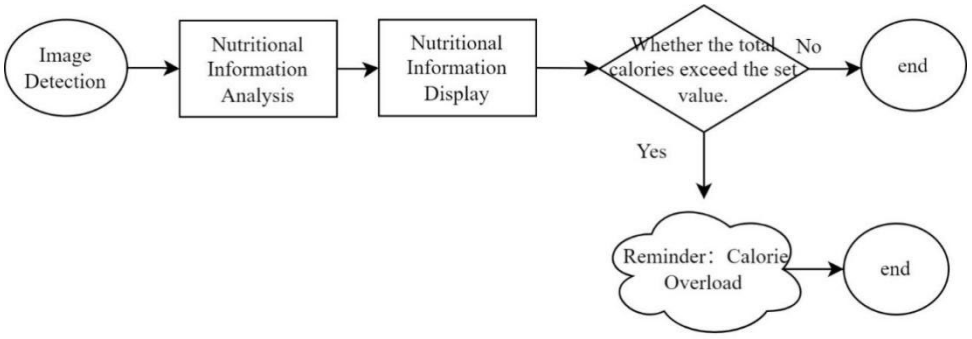

Figure 4: Nutritional assessment module architecture

The nutritional assessment module design is divided into three sub-modules. The first module is the result identification, storage, and analysis module. This module calls the corresponding nutritional component database in the background based on the dish information identified from images, analyzes the nutritional composition by weight, and calculates the total calorie content of the dish represented in the image. The second module is the nutritional information visualization display module, which outputs the pre-set nutritional information of dishes in tabular format. The third module is the nutritional information reminder module, which defines configuration rules for reminder text templates. For example: "Hello, ${user_name}, the calorie intake from ${dish_name} is ${num_calorie}, which has exceeded your expected value." Based on the recognition results from the model, the variables marked

ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

with $ in the configuration template are replaced to generate the reminder text. If the intake does not exceed the user's initially set value, no reminder will be issued.

3. Results

3.1. Model Evaluation

Table 1 presents the results of testing the YOLO v5s model. Overall, the model demonstrates high accuracy, achieving a mean Average Precision (mAP) of 0.97. In terms of prediction accuracy, the class with the highest precision is Shui Zhu Rou Pian (0.99), while the class with the lowest precision is Gong Bao Ji Ding (0.93). The accuracy of the remaining classes is 0.95 or higher.

Regarding recall, Gong Bao Ji Ding has the highest recall value (1.00), while Shui Zhu Rou Pian has the lowest recall value (0.89). For theF1-score, Yu Xiang Rou Si achieves the maximum value (0.98), whereas the minimum value is for Shui Zhu Rou Pian (0.94).

From these metrics, it is evident that the classes performing well overall are Yu Xiang Rou Si and Kuo Shui Huang Gua, while the class that does not perform as well is Shui Zhu Rou Pian.

Table 1: Results of Evaluation

No.	Class	mAP	Precision	Recall	F1-score
1	Yu Xiang Rou Si	0.99	0.98	0.99	0.98
2	Shui Zhu Rou Pian	0.93	0.99	0.89	0.94
3	Mapo Tofu	0.98	0.95	0.97	0.96
4	Qing Jiao Chao Rou	0.97	0.96	0.96	0.96
5	Gong Bao Ji Ding	0.98	0.93	1	0.97
6	Kuo Shui Huang Gua	0.99	0.96	0.98	0.97
Average		**0.97**	**0.96**	**0.97**	**0.96**

Figure 5 illustrates the results of the loss functions for the YOLO v5s model. Overall, all types of loss values decreased to a low range within the first 50 training epochs, indicating that the model performed well during training. Among the losses, Class Loss was initially the highest, reaching around 3.4, while Box Loss was the lowest, at approximately 2.5. However, Class Loss experienced a rapid decline shortly after the training began, stabilizing around 0.4 after 100 epochs, whereas Box Loss stabilized at around 0.5. After 100 epochs of training, the highest loss value was Object Loss, which was about 1.0. This higher Object Loss compared to the other two may be attributed to an imbalanced sample distribution or inaccuracies in some annotations.

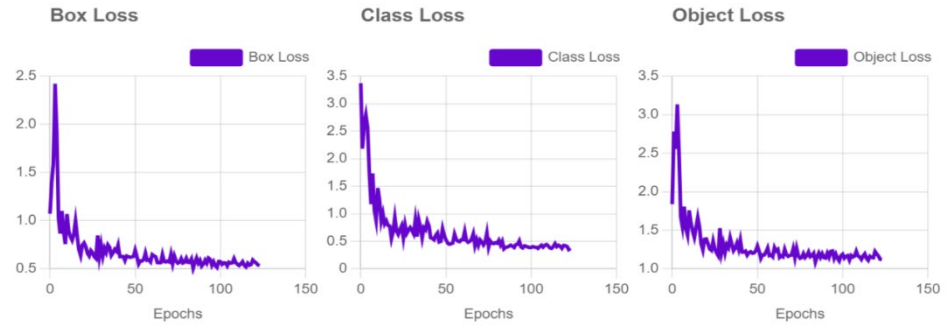

Figure 5: Loss function performance

Figure 6 displays the sample test results for each class, all of which were successfully identified. Table 2 contains the corresponding nutritional information.

ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

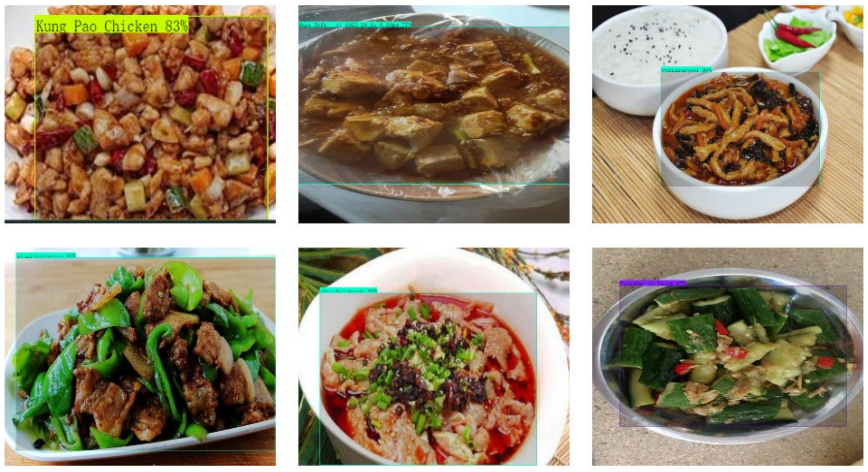

Figure 6: Results of detection

Table 2: Results of nutritional information

No.	Dish	Calorie(kcal)	Protein (g)	Fat(g)	Carbohydrate(g)
1	Yu Xiang Rou Si	263	27.4	12.6	11.2
2	Shui Zhu Rou Pian	466	39.1	26.2	19.7
3	Mapo Tofu	383	30.2	20.8	19.4
4	Qing Jiao Chao Rou	425	25.4	31.0	13.8
5	Gong Bao Ji Ding	360	41.8	15.6	15.8
6	Kuo Shui Huang Gua	126	3.1	7.8	11.4

4. Conclusion

This article designs an integrated system based on the YOLO v5s model that combines image recognition and nutritional assessment. It first introduces the overall framework of the system and the data preprocessing process, including the tools and steps used for data annotation. Next, it discusses the framework of the YOLO v5s model, including the loss function, evaluation metrics, and the design of the nutritional assessment module. The final section presents the detection results of image recognition and the display of nutritional information. From the results of model operation, it can be observed that the model designed in this article performs well, with accurate predictions and identifications, providing a useful reference for researchers in future studies on dish identification. However, on the other hand, this article has certain limitations; based on the performance of the loss function, there may be issues related to insufficient data annotation and imbalanced sample distribution.

Acknowledgements

This work was supported by the Open-End Fund of Key Laboratory of Sichuan Cuisine Artificial Intelligence, Chengdu, Sichuan (Program No. CR23Z10): Deep Learning-based Image Recognition and Nutritional Assessment of Sichuan Cuisine Dishes

References

[1] Zhang Zicheng, Xue Yunlian, Xu Jun. Prevalence and influencing factors of suboptimal health among urban middle-aged and elderly residents in China[J]. Chinese Journal of Public Health, 2023, 39(1): 27-31.

[2] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. arXiv e-prints, 2015: arXiv:1506.02640.

[3] Jiang P, Ergu D, Liu F, et al. A Review of Yolo Algorithm Developments[J]. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021): Developing Global Digital Economy after COVID-19, 2022, 199: 1066-1073.

[4] Yin H, Chen M, Fan W, et al. Efficient Smoke Detection Based on YOLO v5s[J]. Mathematics, 2022,

ISSN 2616-5775 Vol. 7, Issue 9: 11-17, DOI: 10.25236/AJCIS.2024.070902

10(19).
[5] Ryu J, Won D, Lee Y. A Study of Split Learning Model[C]//2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). 2022: 1-4.
[6] Xu M, Yoon S, Fuentes A, et al. A Comprehensive Survey of Image Augmentation Techniques for Deep Learning[J]. Pattern Recognition, 2023, 137: 109347.
[7] Zhao T, Wei X, Yang X. Improved YOLO v5 for Railway PCCS Tiny Defect Detection[C]//2022 14th International Conference on Advanced Computational Intelligence (ICACI). 2022: 85-90.
[8] Du S, Zhang B, Zhang P, et al. An Improved Bounding Box Regression Loss Function Based on CIOU Loss for Multi-scale Object Detection[C]//2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). 2021: 92-98.
[9] Yu S, Zhu F, Chen D, et al. Multiple domain experts collaborative learning: Multi-source domain generalization for person re-identification[J]. arXiv preprint arXiv:2105.12355, 2021.