

# Mahjong detection based on YOLOv5

Jiawei Zhang<sup>1</sup>, Ran Zhi<sup>2</sup>, Ji Zhou<sup>2,3,4</sup>

<sup>1</sup>Changchun College of Electronic Technology, Changchun, Jilin province, 130051, China

<sup>2</sup>SHENZHEN MSU-BIT University, Shenzhen, Guangdong province, 518100, China

<sup>3</sup>1120200283@smbu.edu.cn

<sup>4</sup>Corresponding author

**Abstract.** Mahjong is a captivating game that enjoys popularity in China, Japan, and various other Asian nations. The rapid progress of artificial intelligence technology has spurred significant research attention towards achieving automated mahjong matches. Central to this effort, the precision of mahjong card recognition plays a pivotal role in determining the effectiveness of subsequent tasks. The traditional image recognition method mainly relies on the similarity measurement between the target and the template, whose accuracy and stability cannot meet the needs of the application due to changes in external factors such as the angle and lighting of the mahjong image acquisition in the actual game. In this article, a mahjong intelligent recognition method based on YOLOv5 is proposed. Specifically, this article applies image recognition technology to the AI training of national standard mahjong, and quantitatively analyzes the impact of mahjong image quality (such as shooting angle, mahjong size) on model prediction results. The results show that the proposed method based on YOLOv5 can achieve robust recognition of mahjong, which can bring some new insights to the field AI games.

**Keywords:** Mahjong, Object Detection, YOLOv5, deep learning.

## 1. Introduction

Mahjong, as an ancient and widely popular game, is deeply rooted in Asian culture and also has a large player base around the world. In China, a large number of people participate in mahjong activities, as the joke "1 billion Chinese 900 million play mahjong"[1], which shows the status of mahjong in the eyes of Chinese people. As an indispensable thing in daily life, a large number of mahjong battle platform keep born. To improve the fun of the game, much effort has been devoted to developing diversity functions such as cards recording, cards tips or even automatic battles. To this end, as the key step of the aforementioned functions, Mahjong card recognition has attracted increasing attention in recent years.

The more common mahjong generally has 136 tiles, which can be divided into five categories: 36 'wan' tiles, 36 'bar' tiles, 36 'tube' tiles, and 28 'wind' tiles. A deck of mahjong can have up to 34 patterns, which makes automatic recognition of mahjong tiles a very challenging topic, especially when human-machine hosting is required. Most of the early mahjong recognition methods relied on template matching, that is, calculating the similarity between the target and the template to determine the type of mahjong tiles. However, due to the complexity of actual application scenarios, such as the

uncertainty of mahjong imaging angle, imaging distance or lighting, the above method cannot meet actual application needs. Due to the swift advancements in artificial intelligence technology, notably the groundbreaking achievements of models such as AlphaGo in chess and card games, deep learning-based research in the domain of mahjong recognition has yielded significant and noteworthy outcomes. In the Mahjong world, the most advanced AI is undoubtedly Suphx [2], which is carefully designed based on the rules of the Japanese Mahjong game. Although the Japanese Mahjong rules are similar to the Chinese national standard Mahjong, there are still big differences and the biggest differences is that there are 81 types in national standard Mahjong while only about 40 types in Japanese Mahjong. In addition, the national standard mahjong requirements to at least 8 times to and card. Therefore, the application of Suphx's mahjong model to the national standard of mahjong is unfeasible. Currently, there is no record of anyone successfully completing the AI model training for the national standard mahjong. Apart from the inherent complexity of the national standard mahjong, a critical factor contributing to this challenge is the absence of a suitable dataset for training the AI model. Although there are videos of national standard mahjong games on the Internet, there is a lack of relevant card logs for AI to analyze and learn. Therefore, if you can automatically generate the mahjong card spectrum log through the video of the past games, it will make a considerable contribution to the AI training of the national standard of mahjong.

In the present era, deep learning-based object detection algorithms primarily fall into two categories: One-Stage and Two-Stage methods [3]. Considering the above issues, in this paper, we propose a Mahjong recognition method based on YOLOv5 which is a One-Stage algorithms. Specifically, we first collected and produced a mahjong image recognition dataset to make up for the limitations of insufficient existing data for training. In addition, unlike the aforementioned template matching method, based on the idea of object detection, we introduce the YOLOv5 method to train the mahjong detection model. Finally, we further quantitatively analyzed the mahjong recognition results in different scenarios, such as when the size and imaging angle of mahjong changed significantly.

## 2. Method

### 2.1. Overview

Mahjong is an ancient and popular strategy game, but, in the context of modern technology, automating the images of mahjong tiles becomes an interesting and challenging problem. This section details a mahjong card identification method based on deep learning techniques and how to construct a mahjong card dataset for training and evaluating models.

### 2.2. Data collection and preprocessing

Constructing a high quality mahjong dataset is essential for training and evaluating mahjong recognition models. The CNN-based method usually require amount of data[4]. Here are the key steps in building the mahjong card dataset:

(1) Mahjong tag image acquisition: taking or collecting mahjong tag images. In the current mahjong field, there is no public mahjong data set available, and most of the online pictures are not clearly identified as mahjong faces and used as a data set. And the screen in the mahjong game because the environment is relatively single, in the game, mahjong card and the real mahjong card slightly different, so it can not be used. Therefore, the data set required for the experiment should be obtained by taking the screenshot of the mahjong game video, and each picture contains the hands of the players from the current perspective. It should be noted that, because of the limitation of the shooting Angle, the placement of mahjong is about  $30^\circ$  with the horizontal line. A total of 362 images were obtained through screenshots, and a total of 34 types of labels included 1m-9m (one character to nine characters), 1s-9s (one of bamboo to nine of bamboo), 1p-9p (1 dot to nine dots), dongfeng, nan, xi, bei, bai, fa, zhong (east wind, south wind, west wind, north wind, white dragon, green dragon, red dragon). These images should cover different types of mahjong tiles, various patterns and numbers on different deck faces. Make sure that the various lighting conditions and angles are included. In

addition to manually take some mahjong images, we also get images from the Internet, mahjong game applications or other open source data sets. The specific number of labels for each class is shown in Table 1.

**Table 1.** The specific number of labels for each class

|        |          |     |     |     |     |     |       |     |     |
|--------|----------|-----|-----|-----|-----|-----|-------|-----|-----|
| label  | 1m       | 2m  | 3m  | 4m  | 5m  | 6m  | 7m    | 8m  | 9m  |
| number | 124      | 130 | 169 | 159 | 163 | 141 | 185   | 152 | 122 |
| label  | 1s       | 2s  | 3s  | 4s  | 5s  | 6s  | 7s    | 8s  | 9s  |
| number | 125      | 139 | 135 | 160 | 144 | 137 | 133   | 120 | 95  |
| label  | 1p       | 2p  | 3p  | 4p  | 5p  | 6p  | 7p    | 8p  | 9p  |
| number | 148      | 160 | 184 | 168 | 175 | 162 | 171   | 161 | 160 |
| label  | dongfeng | nan | xi  | bei | bai | fa  | zhong |     |     |
| number | 90       | 77  | 58  | 52  | 55  | 61  | 81    |     |     |

(2) Unified image size: adjust the image to a uniform size so that the model can process images of the same size. Usually, the input image size for the YOLOv5 is  $640 \times 640$  pixels or some other appropriate size.

(3) Data enhancement: The data set is enhanced by random rotation, mirror flipping, brightness and contrast adjustment to improve the robustness of the model.

(4) Create a label file: Create a label file for each mahjong card image, including category labels (e.g., "one barrel") and bounding box information (including the upper left and lower right coordinates).

(5) Dimnotation tools: manually annotate images using annotation tools such as Labellmg or custom tools.

### 2.3. Mahjong card identification

The design of mahjong card recognition method needs to consider not only image processing, but also the organic combination of deep learning technology. We will present an approach based on the Convolutional Neural Network (CNN) and leverage the YOLOv5 (You Only Look Once version 5) algorithm, known for its real-time target detection capabilities and efficient target recognition. Subsequently, our focus shifts to annotating the unique attributes of mahjong, specifically the individual cards. Local features involve directing the network's attention to pivotal local regions and extracting features from these areas. Common techniques for extracting these local features include image segmentation, skeleton key point positioning, and foreground segmentation of pedestrians, among others.

YOLOv5 uses batch normalization (Batch Normalization) in its architecture. YOLOv5 is an object detection algorithm that performs object detection in a deep learning framework, especially for detecting objects in images. "YOLOv5, the latest addition to the YOLO architecture family, boasts impressive features. It excels in detection accuracy, demonstrates swift inference speed, with the potential to achieve a remarkable 140 frames per second for detection. In contrast, the file size of the YOLOv5 target detection network model is notably smaller, nearly 90% less than YOLOv4, making it a suitable choice for real-time deployment on embedded devices. YOLOv5 offers a compelling combination of superior detection accuracy, lightweight design and fast detection speed [6][7]. Batch normalization is used in YOLOv5 for each convolutional layers of the network to help achieve a faster and more stable model training, and improve detection performance [8]. The role of batch normalization is to standardize the input in each layer, which helps to reduce the internal covariate drift and improve the training speed and performance. This is a common technique used to improve the training process of deep neural networks [9]. In YOLOv5, batch normalization helps the network to better adapt to different types of targets and scenarios and improve the accuracy of detection. Batch normalization has the following advantages:

(1) Standardized processing. The core idea of batch normalization is to standardize the input in each layer of the neural network. For each small batch of data, batch normalization calculates the mean and variance of that batch and uses these statistics to normalize each input feature. This helps to keep the input distribution stable for each layer.

(2) Learnable parameters. In addition to standardized processing, batch normalization introduces learnable scaling (scale) and translation (shift) parameters. These parameters allow the network to adaptively learn appropriate linear transformations to better fit the data. Through these parameters, the network can learn to recover a certain representation ability, rather than just performing a linear transformation.

(3) Applied to each layer. Batch normalization is usually applied to each layer of a neural network, including fully connected layers, convolutional layers, and before the activation function. This ensures that each layer in the network benefits from the normalization of the input and the tuning of the learnable parameters.

(4) Solving the internal covariate drift. One of the main effects of batch normalization is to reduce the internal covariate drift, because it keeps the mean and variance of each input stable. This helps to avoid drastic changes in the input distribution of each layer during training, thus reducing the challenges of training, allowing the use of higher learning rates, and improving the training speed.

(5) Accelerated training and regularization effect. Batch normalization not only accelerates the training of neural network, but also has a certain regularization effect, which helps to prevent overfitting.

YOLOv5 is a version of the YOLO series, which simultaneously handles the positioning and classification of targets in an efficient way and is suitable for real-time image processing. The following are the main steps of the algorithm:

(1) Input image preprocessing. First, we need to preprocess the input image and adjust it to the input size of the network, usually  $416 \times 416$  pixels. This step helps to ensure that the network can adapt to images of different sizes.

(2) Network structure. YOLOv5 adopts a deep residual network structure, which contains multiple convolutional layers and residual connections to improve the representation ability of the model.

(3) Target detection. The output of the network includes the category and location information of the target. For mahjong card recognition, we trained the model to be a category that could detect each card, such as "one tube", "thirty thousand", and the position of the cards, usually represented by the bounding box (Bounding Box).

(4) Post-processing. In order to further improve the identification accuracy, we usually use non-maximum suppression (NMS) to remove redundant bounding boxes and ensure that the final output is the accurate position and category of mahjong tiles.

#### *2.4. Model training and fine-tuning*

In the mahjong card recognition task, the training of the model is crucial. We need to prepare an image dataset containing mahjong cards and annotate the correct category and location information for each card. This process may take some time and effort, as the quality of the dataset is very important for the impact of model performance. The model training has the following steps:

(1) Data preparation: Build a data set containing various mahjong card samples. These samples should include the pictures of different types of mahjong tiles, the pictures of the tiles under different lighting conditions, and the images taken at various angles and distances. With diverse data sets, we can increase the robustness of the model.

(2) Annotation data: Mark the correct category and position information for each card. This is usually manual, but annotation tools can also be used to accelerate the process.

(3) Model selection: select the appropriate YOLOv5 model, and fine-tune it according to the task requirements. During the fine-tuning process, the annotated data can be used to train the model to improve its ability to identify mahjong cards.

(4) Model evaluation: During the training process, the performance evaluation of the model is needed. This can be done by retaining a portion of the data for validation or by using cross-validation.

### 3. Experiment

#### 3.1. Experimental settings

The details of our experimental software and hardware are listed in Table 2. During training, YOLOv5s and YOLO5m were selected as weight models, with epochs set to 200 and base-size set to 16. When prediction, the impact of detection error in practical application is far greater than the impact of detection failure, which will greatly increase the difficulty of manual review. Therefore, in order to reduce repeated anchor frames and maintain high precision, IoU threshold was set to 0.3 and confidence threshold to 0.7.

**Table 2.** Details of experimental environment

| parameters | value                   |
|------------|-------------------------|
| GPU        | NVIDIA GeForce RTX 3070 |
| CPU        | Intel Core i5-13600K    |
| System     | Windows11               |
| CUDA       | 11.8                    |
| Python     | 3.8                     |

#### 3.2. Evaluation indicators

The accuracy (Precision, P) and recall (Recall, R) are used to evaluate the performance of the model, which can be calculated by equation (1) and (2).

$$p = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (1)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (2)$$

Where  $N_{TP}$ ,  $N_{FP}$ ,  $N_{FN}$  represents the number of mahjong being correctly identified, not correctly identified, and incorrectly identified in the test results [10].

#### 3.3. Result analysis

In order to verify the effectiveness of this method, we conducted detailed comparative experiments, and the results are shown in Table 3. The time spent to train the model of YOLOv5s is about 2h and the accuracy can reached 93.29%, while the recall R is 90.01%. Considering only the frontal orientation and severe occlusion, there were 694 tiles in the test set, among which 556 were successfully identified, 114 were not identified as mahjong, and 5 were incorrect. The accuracy rate P was 99.11%, and the recall rate R was 80.16%. In contrast, the time spent on model training in YOLO5m reached 4h, and the value of model accuracy P was 98.87% and R is 98.21%. Similarly, only identifying significance was considered. Among the 694 tiles in the test set, 675 were successfully identified, 6 were not identified, 13 were incorrect, P was 98.11% and R was 97.26%.

**Table 3.** Performance comparison of different methods

| algorithm | P/%    | R/%    | time/h |
|-----------|--------|--------|--------|
| yolov5s   | 99.11% | 80.16% | 2      |
| yolov5m   | 98.11% | 97.26% | 4      |

By comparing the time and accuracy required for the training of the two models, the following conclusions can be obtained. In the field of mahjong identification, there is almost no large difference between the same labels, the form is relatively single, and the main factors affecting the success rate of

identification are the Angle and placement direction. In the case of the limited mahjong placement Angle used by the data set is relatively single, the use of YOLO5m model can get a better identification effect. With the same recognition accuracy, the recall rate of YOLO5m is 21.3% compared with yolov5s.

#### 4. Discussion

Through the above research method, a targeted model that can monitor the situation of the game in real time. Analyze each card in the game and provide key information such as type, number and location. In addition, it is expected to reduce the workload of collecting data sets and improve the accuracy of the model. Through this study, we will make an important contribution to the automation and intelligent development of mahjong games in China, and provide a valuable reference for the research in related fields. At the same time, in the future, the research can also be extended to other card games, board games and other applications, to improve the level of automation and intelligence.

Of course, there are still some shortcomings in the current study. For example, our dataset is mainly focused on specific type, mahjong tiles, considering expanding the dataset to cover more changes to cope with different styles in the future. In addition, interacting with real-world mahjong tiles may involve more environmental pollution and complexity, such as camera perspective, and ambient light, requiring more research to adapt to these challenges.

#### 5. Conclusion

Automatic recognition of mahjong is crucial to improving the fun of mahjong games and expanding the application of artificial intelligence in the gaming field. In response to the challenges of insufficient available mahjong training data sets and poor recognition accuracy based on template matching methods in existing research, this paper proposes a mahjong detection and recognition model based on YOLOv5. Specifically, we collected and preprocessed a mahjong image dataset, which was further manually standardized for model training. We compared the recognition accuracy of different YOLOv5 methods for different mahjong, and the results can reach 90%. We finally discuss the impact of the angle of mahjong image imaging on the model recognition results and give possible solutions for correction. A large number of experiments have verified the effectiveness of this research, which can bring new insights to AI Mahjong.

#### Authors contribution

All authors contributed equally to this research, and their names are listed in alphabetical order.

#### References

- [1] Ding Wei. Mahjong Fate Record [ J ]. Wen Shi Tian Di, 2022 ( 12 ) : 71-74.
- [2] Junjie Li, Sotetsu Koyamada, QiweiYe, Guoqing Liu, Chao Wang, Ruihan Yang, Li Zhao, Tao Qin, Tie-Yan Liu, Hsiao-Wuen Hon. Suphx: Mastering Mahjong with Deep. Reinforcement Learning[m], 2003, 1-2
- [3] Cheng Zehua, Han Junying. Lightweight Improvement of YOLOv5 Mask Wearing Detection Algorithm. Software Guide:1-6 [2023-11-08].
- [4] Zhifang Wu, Jiuqiang Han, Erhu Liu, and Hongqiang Lyu. 2021. A Human-Robot Interactive Mahjong Playing System Based on Visual Recognition Using a Convolutional Neural Network. In Proceedings of the 5th International Conference on Computer Science and Application Engineering (CSAE '21). Association for Computing Machinery, New York, NY, USA, Article 63, 1-8.
- [5] LUO Hao, JIANG Wei, FAN Xing and ZHANG Si-Peng. A Survey on Deep Learning Based Person Re-identification. ACTA AUTOMATICA SINICA, Vol. 45, pp: 2032-2049,2019.
- [6] Bin Yan, Pan Fan, Xiaoyan Lei, Zhijie Liu and Fuzeng Yang. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. Remote Sensing, vol. 13(9),2021.

- [7] Zhaoyi Chen , Ruhui Wu , Yiyang Lin , Chuyu Li , Siyu Chen , Zhineng Yuan , Shiwei Chen and Xiangjun Zou. Plant Disease Recognition Model Based on Improved YOLOv5. *Agronomy*, Vol. 12(2),2022.
- [8] Zhifang Wu, Jiuqiang Han, Erhu Liu and Hongqiang Lyu. A Human-Robot Interactive Mahjong playing System Based on Visual Recognition Using a Convolutional Neural Network. *Proceedings of the 5th International Conference on Computer Science and Application Engineering*, Vol. 63, pp:1-8,2021.
- [9] YangHongyun, WanYing, WangYinglong and LuoJianjun. Identification of Rice Diseases Based on Batch Normalization and AlexNet Network. *Progress in Laser and Optoelectronics*, Vol.58, No.6,2021.
- [10] Zhou Ailing, Tan Guangxing. Traffic light detection algorithm based on YOLOv5s [J]. *Journal of Guangxi University of Science and Technology*,2023,34(04):69-76.