Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **10**, Number **2**, May **2025**

Recognition of the Oracle-bone Inscriptions Based on Deep Learning

Ke-Ke Sun

School of Design Art
Xiamen University of Technology, Xiamen 361000, China
sunkeke@xmut.edu.cn

Miao-Miao Kang*

School of Design Art
Xiamen University of Technology, Xiamen 361000, China
2322191017@s.xmut.edu.cn

Tian-You Xue

School of Automotive and Mechanical Engineering
Xiamen University of Technology, Xiamen 361024, China
2421011038@stu.xmut.edu.cn

Mu-Hsuan Huang

Department of Library and Information Science
National Taiwan University, New Taipei City 237303, Taiwan (R.O.C)
mhhuang@ntu.edu.tw

*Corresponding author: Miaomiao Kang

Received November 8, 2024; revised February 18, 2025; accepted April 1, 2025.

Abstract. *Oracle-bone Inscriptions, the earliest writing system in China, is the source and cultural root of Chinese characters. It occupies an important position in Chinese civilization and human civilization. Its identification work can provide important clues for the study of ancient history, which is of great significance meaning. In this study, the field investigation and image collection were carried out in the Yin Ruins which has the earliest and most abundant corpus of Oracle-bone Inscriptions found in China. Following data enhancement processing, a series of target detective and recognition experiments were conducted by using YOLOv8 model combined with optimized Backbone and Head architectures and the anchorless Split Ultralytics Head. Comparing three different YOLOv8 models, the experimental results show that the YOLOv8s model leads the other two models with a mAP value of 93.5%which is 3.2 percentage point higher than YOLOv8x and 2.2 percentage point higher than YOLOv8n. The YOLOv8s model achieves the highest mean average precision while maintaining a compact model size, demonstrating an optimal balance between performance and efficiency.*
Keywords: Oracle-bone Inscriptions, Deep learning, Recognition

1. Introduction.

Deep learning technology improves the Oracle-bone Inscriptions identification method. Fu et al. [1] created the OBI-100 dataset with 100 categories and more than 128,000 samples. By improving the model structure of LeNet, AlexNet, and VGGNet, their recognition and classification accuracy of Oracle-bone Inscriptions reached 99.5%. Yuan et al. [2] proposed a recursive graph neural network (R-GNN) for automatic oracle bone inscriptions

font recognition, combining CNN, GNN, and RNN to extract local features and global context information. They achieved a Top-1 accuracy of 88.2% and an F1-score of 88.0%. In font retrieval, the Top-1 accuracy reached 98.0% and the mAP was 70.5%, demonstrating strong robustness and classification performance. Lin et al. [3] utilized the Maximally Stable Extremal Regions (MSER) algorithm in conjunction with a deep learning network to improve the accuracy of Oracle-bone Inscriptions radical detection and recognition to 94.2%. The detection accuracy is improved by 24%, and the Batch Normalization (BN) and Dropout layers improve the Top1 accuracy by 6.9%. Li et al. [4] based on generative adversarial network (GAN) framework to enhance the Oracle-bone Inscriptions in the defined problem category, by generating synthetic data using a generator. The accuracy of long-tailed Oracle-bone Inscriptions recognition is improved by an average of 8% within this framework. Guo et al. [5] proposed a novel methodology for unsupervised domain adaptation aimed at transferring labeled handwritten Oracle-bone Inscriptions to unlabeled scanned data. By improving the Inception-v3 model and introducing a convolutional attention module, feature extraction is enhanced. The experimental results indicate that the model consistently exhibits outstanding recognition performance, even when faced with blurred, occluded, and mutilated characters. Wang et al. [6] proposed an unsupervised discriminative consistency network for Oracle-bone Inscriptions recognition. The model's consistency is enhanced to improve its robustness against wear, stains, and distortion. A 15.1% improvement in performance was observed when compared to structured texture separation networks for Oracle-bone Inscriptions scanned data. Gao et al. [7] put forth a methodology for image translation based on generative adversarial networks, whereby Oracle-bone Inscriptions images are translated into modern Chinese characters. A symmetric structure is designed using encoders and decoders to facilitate the translation of images from Oracle-bone Inscriptions to modern Chinese characters. The accuracy of the training set retrieval is enhanced from 90.8% for top10 to 99.4% for top 200. Guo et al. [8] proposed a texture mapping-based Oracle-bone Inscriptions extractor designed to extract Oracle-bone Inscriptions from three-dimensional models. The use of symmetry points in the texture mapping process facilitates the extraction of 3D Oracle-bone Inscriptions from Wavefront files containing 3D Oracle-bone Inscriptions models. The feasibility of the algorithm framework for extracting scratches from 3D Oracle-bone Inscriptions is verified by experiments. Yue et al. [9] put forth a novel deep learning-based approach, which solved the problem of data imbalance in the Oracle-bone Inscriptions recognition by modifying Generative Adversarial Network (GAN) to enhance the original Oracle-bone Inscriptions dataset. Furthermore, they proposed a novel deep learning model, designated as C-A Net, which can effectively mitigate the overfitting issue while maintaining a commendable recognition accuracy of 91.10%. Shi et al. [10] proposed an ant colony algorithm-based Oracle-bone Inscriptions recognition algorithm, which improved the efficiency of recommendation systems by expanding the data in the Oracle-bone Inscriptions interest network. Using the one-to-one correspondence between characters and translations in Oracle-bone Inscriptions rubbings, Oracle-bone Inscriptions can be recognized efficiently. Experimental evidence indicates that the recognition rate is nearly 100% for certain specific Oracle-bone Inscriptions topographies. To sum up, recent studies have demonstrated a variety of approaches to enhance the accuracy of Oracle-bone Inscriptions character recognition, including optimization of deep learning models, data enhancement, and the design of novel network structures. These methods significantly improve the recognition accuracy, improve the performance of both the convolutional neural network (CNN) model and the recurrent graph neural network, and make the model have better generalization ability and robustness. However, there are also demands for

large-scale datasets, risk of over fitting, and challenge for long-tailed distributed data processing.

Extending to the whole field of character recognition research, deep learning techniques has also played an important role. Hazra et al. [11] proposed a convolutional neural network (CNN) architecture for recognizing handwritten Bengali and Maith Majek characters. Tests were conducted on three publicly available Bengali datasets: cMATERdb, ISI Bangla, and the newly created Mayek character dataset "Mayek27". The experimental results indicate that a recognition accuracy of 99.27% was achieved on the "Mayek27" dataset. VERMA et al. [12] proposed an algorithmic approach that utilizes region recognition to address the problem of online handwritten Gurmukhi character recognition. A support vector machine model is used to recognize the strokes within each region. The results of the experiments demonstrate that the accuracy of region recognition is 95.3%, while the accuracy of character recognition is 74.8% on the 428-character dataset. Hamplová et al. [13] proposed using the YOLOv8 and Roboflow 3.0 algorithms for segmenting both multi-class and single-class instances of Palmyra Aramaic inscription characters, comparing the performance differences between the two. Label the special datasets to train the model, evaluate the effect with quantitative indicators. The experimental results show that under the multi-class model, the total character accuracy of YOLOv8 reaches 93.5%, whereas the accuracy for single-class instance segmentation is 77.3%. Guo et al. [14] studied Oracle-bone Inscriptions recognition and general sketch recognition. A new hierarchical representation method is proposed, which combines the low-level representation of Gabor filters and the middle-level representation based on sparse encoders. The experimental results indicate that the method achieves a recognition accuracy of 56.0% on the Sketch-20K dataset. Gao et al. [15] proposed a novel dual-stream convolutional neural network to address the challenges of segmenting and recognizing ancient Oracle-bone Inscriptions fragments. The UNet++ and Faster R-CNN architectures have been improved by incorporating residual blocks and bilinear interpolation. Additionally, the article builds a dataset that includes Oracle's segmentation and detection labels, achieving a material classification accuracy of 91.8%. Nabi et al. [16] proposed a gender classification system using deep learning techniques for Urdu handwritten characters. The study utilized a self-constructed dataset that included 400 individuals, evenly split between 50% male and 50% female, resulting in a total of 284,000 samples. A ratio of 7:3 was employed to partition the training and testing sets, resulting in an accuracy rate of 99.63%. Kumar et al. [17] proposed a framework for recognizing handwritten writers based on pre-segmented Gurmukhi characters. This framework incorporates two distinct feature extraction techniques along with two classifiers: k-nearest neighbors (k-NN) and support vector machines (SVM). Experimental results show that the recognition accuracy can reach 89.85% by using support vector machine (SVM) classifier combined with various features, and the accuracy can reach 94.76% under 10-fold cross-validation. Cheng et al. [18] put forward a new graphic descriptor GIST-SC based on the rapid recognition architecture of retrieval, which is used to retrieve calligraphic character images. Experimental evidence indicates that the GIST-SC descriptor and recognition architecture are effective and efficient for the recognition of Chinese calligraphic characters with calligraphy character dictionary. Gao et al. [19] proposed a new framework LUC in the field of oracle bone inscription retrieval. By constructing a multimodal knowledge graph and a domain-aware embedding module, LUC achieves effective retrieval of unknown characters. Experiments show that LUC achieves a mAP of 60.8% on the ReOBI dataset, surpassing EGFF's 47.1%, Trans ReID's 48.4%, and ViT-Baseline's 49%, outperforming other classic methods. To sum up, the above research presented various character recognition techniques, which generally use the convolutional neural network (CNN) framework and

emphasize the importance of feature extraction. These methods improve the accuracy and speed of recognition; however, the disadvantage lies in the high demand for large datasets and the complexity of model training.

Our laboratory has made impactful research outcomes in the field of machine learning and deep learning. Chiu et al. [20] designed an automated garbage recycling system that employs a machine vision and robotic arm control system. This system utilizes the improved YOLOv5 algorithm to achieve high accuracy in recognition. By optimizing the model using the channel pruning technique, the number of parameters was reduced to 2,828,675. This optimization led to a decrease in model size of approximately 60%, while preserving high recognition accuracy. The results of the experiments demonstrated that the success rates for the red, blue, and green bins were 80%, 85%, and 75%, respectively. Gu et al. [21] employed the YOLOv5 algorithm for target detection, improved the model's robustness and generalization by utilizing data enhancement and incorporating a Convolutional Block Attention Module (CBAM). After 75 training cycles, the model achieved a mean accuracy of 99.2%, outperforming other algorithms such as SD, EfficientDet, and Faster R-CNN. The author has achieved substantial progress in anthropometric data measurement and deep learning. Initially, Sun et al. [22] assessed the effectiveness of two algorithms—thin-plate spline interpolation (TPS) and linear interpolation—to address the issue of incomplete anthropometric data for elderly men in Fujian Province, China. The predictive efficacy of the two algorithms was evaluated under various missing data scenarios, using a total of 54 datasets. The findings indicate that the TPS algorithm exhibits minimal prediction error and performs better when faced with a higher proportion of missing data. Subsequently, Sun et al. [23] proposed a weighted regression methodology to rectify incomplete data in an anthropometric database of elderly women. The least squares and weighted regression methods were utilized to analyze the optimization settings and fitting effects by applying data classification and correlation analysis. The findings indicate that the weighted regression approach is more effective than the least squares method for handling data with varying levels of missingness. In a recent study, Sun et al. [24] explored the potential of deep learning for classifying and recognizing Nantong blueprint patterns. The recognition rate is improved by optimizing the SSD model. A comparative analysis is conducted with the Faster R-CNN and YOLOv8 models. The findings indicate that the SSD model achieves the highest accuracy, demonstrating a recognition rate of 79.42% for blue printed fabric.

In summary, recent advancements in the recognition of Oracle-bone Inscriptions character and other ancient characters primarily utilize deep learning algorithms. Current research focuses on enhanced neural networks, recursive graph neural networks (R-GNNs), and generative adversarial networks (GANs). Nonetheless, YOLOv8, the latest algorithm in the field of object detection, has not been widely used for the identification of Oracle-bone Inscriptions and other characters. Therefore, supported by the deep learning technology in the laboratory, this paper chooses Oracle-bone Inscriptions as the research object, and uses YOLOv8 series model to identify Oracle Bone Inscriptions. Through the target detection and recognition experiments, three different models of YOLOv8 series are compared and comprehensive evaluated.

2. **Materials and Methods.**

2.1. **Target object–Oracle-bone Inscriptions.**

The Oracle-bone Inscriptions is the oldest recognizable writing system in China and is also the oldest and most mature form of writing in Chinese history. Oracle-bone Inscriptions primarily refer to those in Yin Ruins, engraved on tortoise shells and animal

bones during the Yin and Shang dynasties, as shown in Figure 1, for divination and event recording.

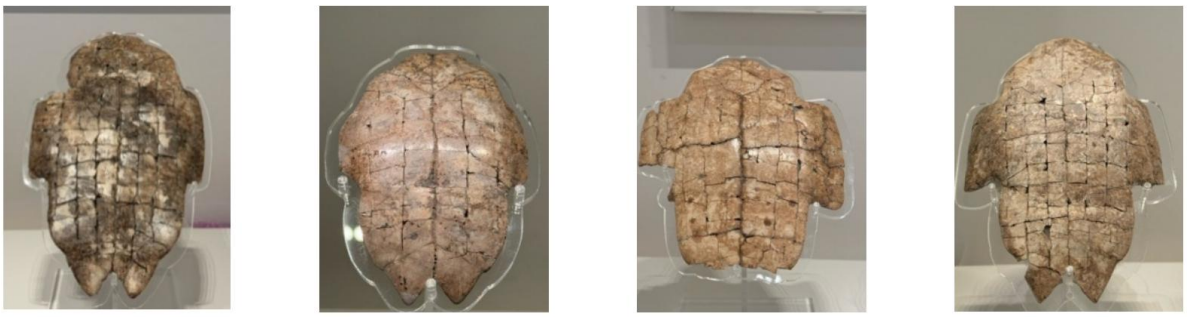


FIGURE 1. Oracle-bone Inscriptions Fragments

Different Oracle-bone Inscriptions characters have different font characteristics, and throughout their evolution, many of these characters have diverged significantly from modern writing forms. To ensure the completeness of the selected text categories and to prevent the model from overfitting by focusing solely on complex or special glyphs during training, the dataset was carefully curated. This paper takes 'bing', 'jia', 'geng', 'zi', and 'yu' (as shown in Figure 2) as key research objects, that is, it contains both glyphs that are distant from modern Chinese characters, such as 'jia' (Figure 2(b)), and involves 'yu' (Figure 2(e)) that is not much different from modern Chinese characters.

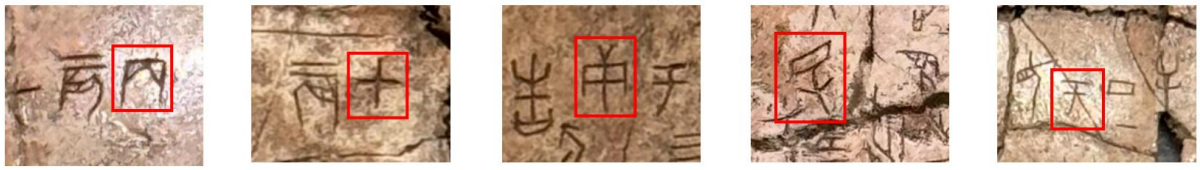


FIGURE 2. Character for Oracle-bone Inscriptions

The character 'bing', shown in Figure 2(a), is a sophisticated and expressive image that vividly represents the rock on which the flag was planted in ancient times. Its ingenious hollow design not only reduces the weight, but also enhances the ease of handling. The character for 'jia' illustrated in Figure 2(b), effectively depicts the moment when a plant's shell cracks open as it begins to sprout. This imagery symbolizes the connection between pieces of armor and reflects the ancient people's reverence for nature and the military. The character 'geng' illustrated in Figure 2(c), represents a man drawing water from a bucket. This imagery symbolizes acquisition and renewal, reflecting the ancient people's harmonious relationship with nature. The character 'zi' depicted in Figure 2(d) features simple, graphic lines that outline the form of a newborn baby. This representation reflects the ancient people's reverence for and celebration of the beginning of life. The character 'yu' shown in Figure 2(e) graphically represents the outline of an ancient musical instrument.

2.2. **Establishment of the dataset.**

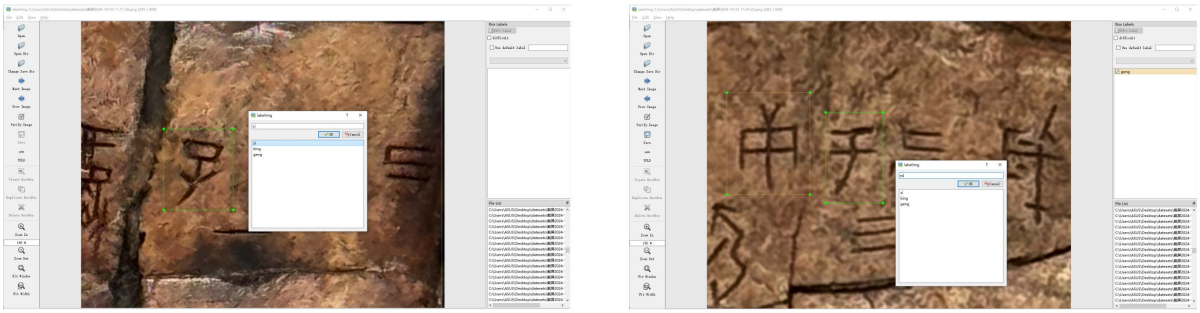
2.2.1 Oracle-bone Inscriptions dataset

The Oracle-bone Inscriptions image data presented in this paper were collected from August 11 to 15, 2024, at the Yin Ruins Museum in Anyang City, Henan Province, the

birthplace of Oracle-bone Inscriptions. The device used for data acquisition is an iPhone 15 Pro. The lighting inside the museum is usually adjusted to the most suitable state for the protection of cultural relics to avoid direct glare and protect Oracle-bone Inscriptions from damage. Therefore, when taking pictures, various shooting angles and distances were employed to comprehensively document the morphological characteristics of the Oracle-bone Inscriptions. This approach ensured the diversity and completeness of the visual data.

2.2.2 Dataset Labelling.

The training images were manually labeled using LabelImg software, as shown in Figure 3. The specific process is as follows: First, start LabelImg, click the 'Open Dir' button in the interface, and select the folder for storing Oracle-bone Inscriptions pictures. Click the 'Change Save Dir' button to choose the storage location of the tags. Next, click the 'Create Rect-Box' button in the toolbar on the left of the interface. This action will draw a rectangular box on the image with the mouse. Once the rectangular box is drawn, the label name will be entered in the pop-up dialog box. Finally, click the 'OK' button to save the annotation file. Once the annotation process is complete, a TXT tag file is generated that contains the coordinates of the center point of the Oracle-bone Inscriptions text target, along with information about the width and height of the annotation box. Subsequently, the collected Oracle-bone Inscriptions target dataset is divided into a training set and a validation set.



(a) (b)

FIGURE 3. The labeling process of the dataset

3. Algorithm.

3.1. YOLOv8.

Most traditional target detection algorithms separate the detection and classification processes into multiple stages, which often leads to slower inference times, making it difficult to achieve an optimal balance between accuracy and speed. In contrast, the YOLO algorithm family's core concept approaches target detection as a regression problem, employing an end-to-end processing methodology. As indicated by its full name, "You Only Look Once," the YOLO algorithm can generate multiple prediction boxes and their corresponding categories simultaneously by executing a single test on an individual image. Furthermore, the YOLO series of algorithms is characterized by a simple network architecture that allows for image processing from a global perspective. The efficiency of the training and inference processes is superior to that of traditional image recognition algorithms. Most importantly, unlike thetraditional target detection algorithms, which detects objects first and then classifies them, YOLO series algorithms can complete the task of target detection and classification at the same time. Considering that the application background of this paper is to the identify Oracle-bone Inscriptions characters in a

picture, different characters may appear in the same picture, the selected algorithm must be capable of identifying the regions where these characters are present simultaneously and of classifying them correctly.Therefore, we select the YOLO series of algorithms to recognize Oracle-bone Inscriptions characters in this paper

YOLOv8 is a target detection algorithm developed by Ultralytics. It has been improved and optimized compared to its predecessor, resulting in greater accuracy and reduced processing time. The core features of the system include advanced Backbone and Neck architectures, an anchorless Split Ultralytics Head, an optimized balance of accuracy and speed, and various pre-training models. The network structure consists of three main components: an input layer, a backbone network, and a header network. Mosaic data enhancement is applied during the input terminal, and the DarkNet53 structure serves as the backbone network. Additionally, the C3 module is replaced with the C2f module, which improves the efficiency of feature extraction. The head network achieves feature fusion and target detection through the use of PANet and a prediction head. PANet combines the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to enhance the network's ability to detect targets at multiple scales. The predictive head outputs classification and coordinate information tailored to the specific requirements of various inspection tasks. The network framework diagram is shown in Figure 4. Among them, CBS refers to the convolution, normalization, and activation functions; SPPF denotes the spatial pyramid pooling module; C2f-PConv module is C2F module with partial convolution; Fusion module is a rapid, normalized module; Up-sample refers to the up-sampling module; Detect denotes the detection header; MaxPool2d indicates maximum pooling; Conv signifies convolution; and Contact represents the feature coupling module; Box. Loss and Cls. Loss refers to the bounding box loss and classification loss, respectively

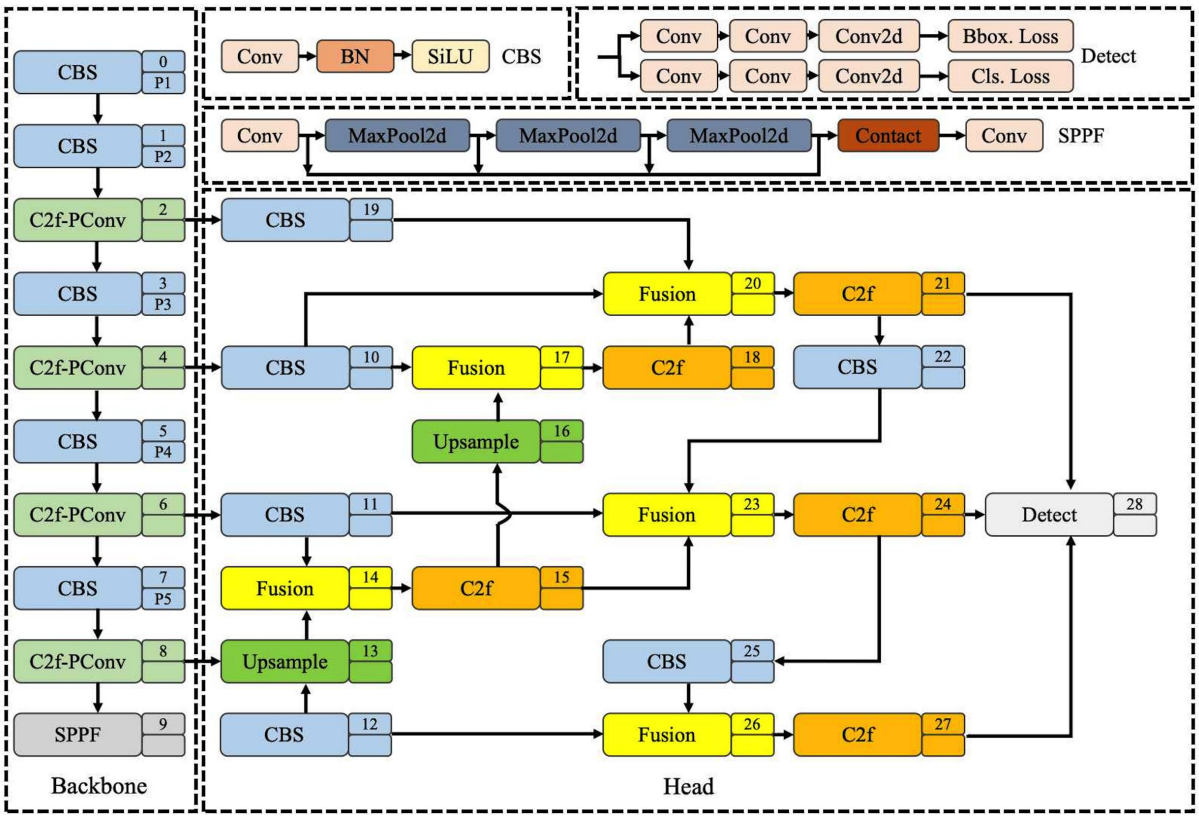


FIGURE 4. Structure Diagram of the YOLOv8 Network

To calculate the loss, the YOLOv8 model employs a positive and negative sample matching strategy within Task-aligned One-stage Object Detection (TOOD). The prediction box score t is calculated as demonstrated in Equation (1):

$$t = S^{\alpha} + \mu^{\beta} \tag{1}$$

where S refers to the classification prediction score, u represents the intersection and concurrency ratio of the prediction frame relative to the true box, and α and β are weighted hyper parameters. The method selects positive and negative samples according to the magnitude of the t-value, thereby enhancing the interaction between classification and localization tasks.

When calculating the regression loss, the Distribution Focal Loss (DFL) is incorporated into the original C-type Intersection over Union (CIOU) loss. This approach directs the network's focus toward the vicinity of the target location. The DFL is calculated as demonstrated in Equation (2):

$$DFL(S_i, S_i + 1) = -(y_i + 1, -y_i) + (y-, y_i)lg(S_i + 1) \tag{2}$$

where y is the true label, y_i, y_i+1 are the two quality labels in the neighbourhood of the true label, and S_i, $S_i + 1$ are the model sigmoid outputs corresponding to the quality labels.

3.2. **YOLOv8 Series Algorithm.**

As shown in Table 1, the YOLOv8 family of algorithms comprises five variants categorized by model size, computational requirements, inference speed, and detection accuracy. The five categories are YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large). YOLOv8m and YOLOv8l are primarily utilized for tasks requiring high-accuracy text recognition. These applications include complex document analysis, text recognition in video frames, autonomous driving, and medical imaging. Therefore, this paper does not compare the two algorithms; instead, it focuses on three specifications of the YOLOv8 model——YOLOv8n, YOLOv8s, and YOLOv8x——for a side-by-side comparison. The YOLOv8n model is designed to operate in memory-constrained environments and is compatible with small hardware, making it well-suited for the experimental environment in this paper. The YOLOv8s model offers greater accuracy than YOLOv8n,particularly in fast-response scenarios, thus achieving a better balance between speed and accuracy. By increasing the network depth and the number of parameters, the YOLOv8x model demonstrates improved accuracy in managing complex target detection tasks, especially in scenes with complex details, detection of small objects or high background interference. Combined with the data set selected in this paper, Oracle-bone Inscriptions-bone Inscriptions characters and backgrounds blend together, this model can be more competitive in detection capability, sensitivity to small objects, flexibility, and lightweight. The selection of these three variant algorithms not only includes the high real-time requirement in the environment of low hardware resources to realize the task of fast text detection, but also includes the detection algorithm with the highest accuracy, which can detect tiny characters in high-resolution pictures, and can compare the applicability and advancement of YOLOv8 algorithm in the field of Oracle-bone Inscriptions character recognition from different angles.

Table 1. Variants of YOLOv8

Variants of YOLOv8	Number of parameterser	Calculation	Inference speed	Precision	Usage scenarios
YOLOv8n	minimum	least	fastest	lowest	Embedded devices, real-time detection
YOLOv8s	small	less	faster	lower	Mobile devices, lightweight applications
YOLOv8m	medium	medium	medium	medium	Autonomous driving, surveillance
YOLOv8l	big	more	slower	higher	Medical imaging, altitude mission
YOLOv8x	biggest	most	slower	highest	Refined object detection, complex scenes

4. **Experiment.**

4.1. **Setting up environment.**

The hardware configuration utilized in this experiment: NVIDIA GeForce RTX 3060 graphics card, 12 GB RAM, and a 512 GB SSD. The software operates on the Windows 10 Professional operating system, and all programs are trained using the deep learning framework Torch 1.13.1 with CUDA 11.6 to enhance operational efficiency. The training parameters are presented in Table 2.

Table 2. Training parameters

Training parameters	Value
Image size	640
Batch size	10
Optimizer	AdamW
Optimizer momentum	0.937
epochs	100
IoU	0.7
Ir0	0.01
Irf	0.01
weight-decary	0.005
CLS	0.5

4.2. **Indicators for experimental evaluation.**

To compare the performance of different models, the evaluation metrics employed include precision (P), recall (R), average precision (AP), mean average precision (mAP), memory usage, and computation and inference time. These metrics facilitate a quantitative assessment of the recognition performance of each model.

Accuracy measures the proportion of positive instances correctly predicted by the model among all instances predicted as positive. In target detection, a high accuracy indicates a low false alarm rate for the model. The calculation formula is provided in Equation (3):

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{3}$$

Recall measures the proportion of correctly recognized positive instances identified by the model compared to all actual positive instances. In target recognition, a high recall indicates a low leakage rate for the model. The calculation formula is provided in Equation (4):

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{4}$$

AP and mAP are metrics used to evaluate the performance of target detection models. They reflect the average detection accuracy across images of varying categories and difficulty levels. The calculation equations are provided in Equations (5) and (6):

$$AP - \int_0^1 P \cdot dR \tag{5}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{6}$$

where TP denotes the number of positive samples correctly identified by the model, FP represents the number of positive samples not recognized by the model, FN indicates the number of negative samples incorrectly identified as positive by the model, and N is the number of categories as 2.

4.3. **Experimental data processing.**

During the training phase, the images undergo preliminary processing using the HSV color model. The dataset's training images are flexibly adjusted based on three parameters: hue, saturation, and value.Secondly, image rotation (by specifying an angle), translation (moving horizontally or vertically in the image plane), scaling (resizing the image), and cropping (applying a skew transformation) are introduced to enhance the dataset. Figure 5 illustrates the training process of the dataset after data enhancement.

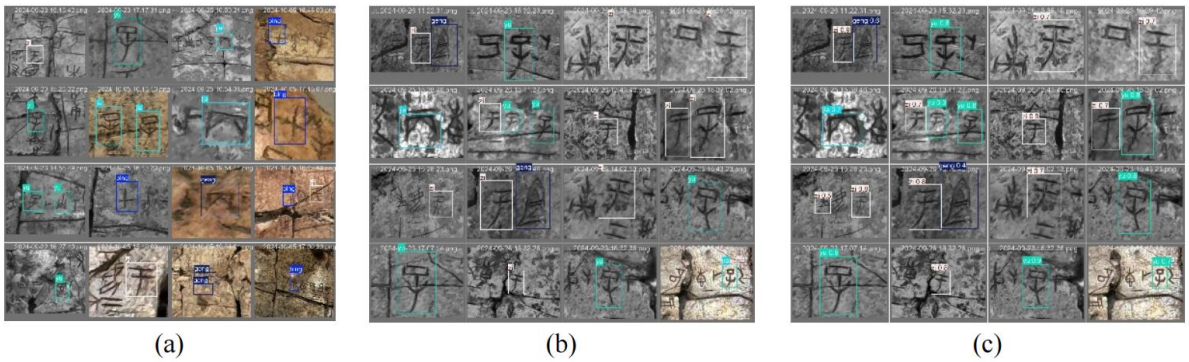


FIGURE 5. Training process image enhancement display

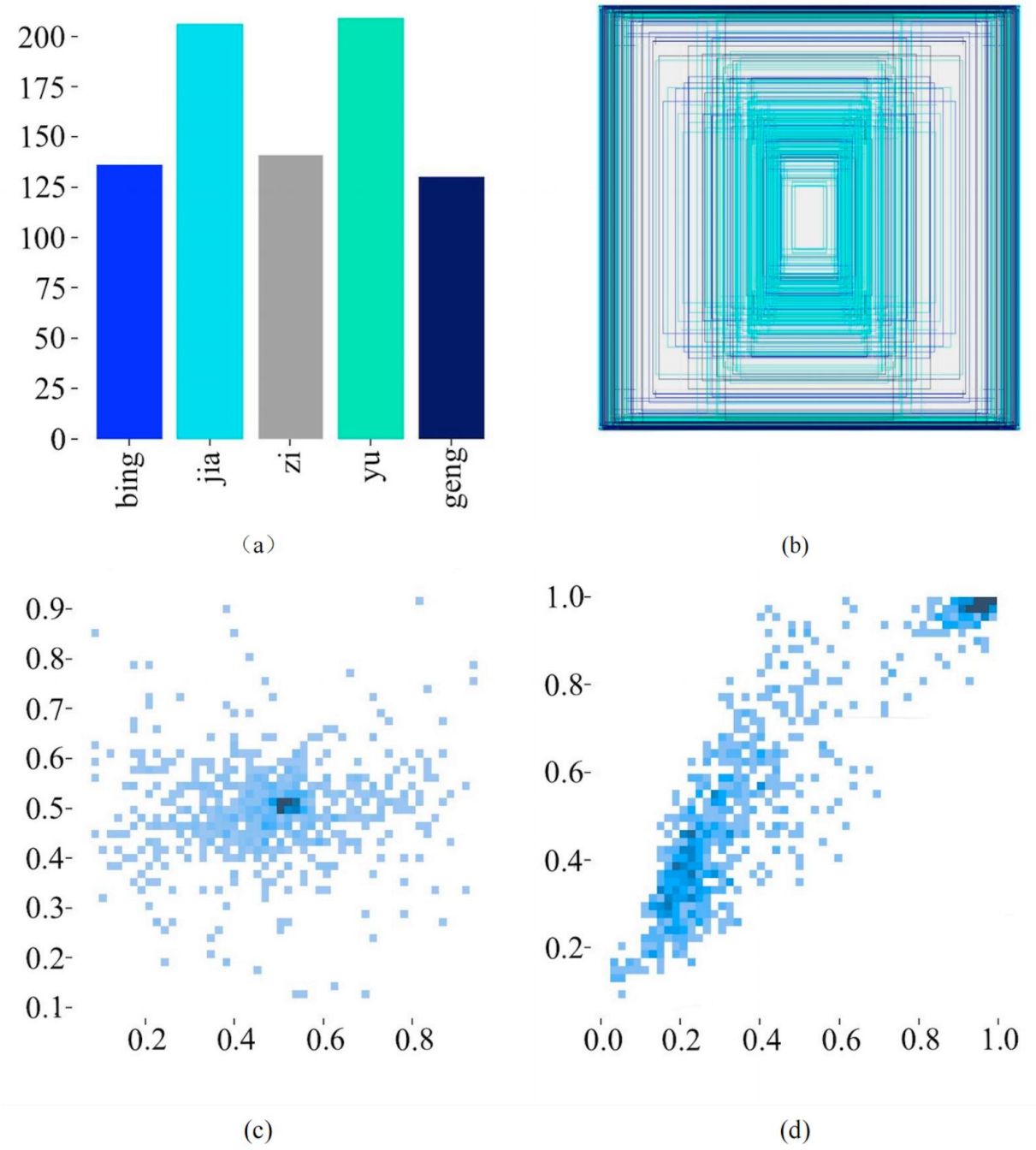


FIGURE 6. Training process image enhancement display

The dataset presented in Figure 6 consists of five Oracle-bone Inscriptions image samples. Figure 6(a) shows the data distribution and its characteristics. Among them, the category 'yu' has the largest number of instances, followed by 'zi', while 'bing' and 'jia' have fewer instances, and 'gene' has the least. This uneven distribution may result from dataset imbalance and sampling bias. As shown in Figure 6(b), the outlines of multiple concentric rectangles or squares are displayed. These outlines represent the iterative process and convergence path of the image enhancement algorithm at different scales. This demonstrates that the algorithm refines image features over several iterations, ultimately reaching a stable state. The prediction boxes for the five font recognition targets are mainly concentrated in the center of the image, with some also appearing at the edges. The boxes tend to cluster toward the center, with only a small portion of samples located

near the image's edges. Figure 6(c) presents information regarding the prediction box. Both the horizontal and vertical axes are marked with numerical range (0 - 1), illustrating the distribution of image features in two-dimensional space. The dense scattered points are concentrated in the central region, indicating that most feature values, after normalization, are clustered within a specific range and may exhibit strong linear correlations. The relative coordinate range in this figure is within (0.3 - 0.6, 0.3 - 0.6). This feature may impact subsequent image classification and recognition tasks,as smaller object sizes could increase the difficulty of recognition and classification. In processing this type of data, special technical measures may need to be taken to improve recognition accuracy. Figure 6(d) illustrates the dimensions of the labeled box, namely its height and width. It also illustrates the distribution of image sizes in two-dimensional space. The dense scatter points are concentrated in the central area, suggesting that the sizes of most images fall within a specific range. This also indicates a strong linear correlation between the width and height of the images. It can be observed that these samples are mainly distributed within the region defined by relative coordinates ranging from (0 to 0.6, 0 to 0.4), and a small number of samples are found individually in the region between (0.8 and 1.0). The overall data indicates that most Oracle-bone Inscriptions samples have a relatively small object size compared to the overall image.

4.4. **Experimental recognition results.**

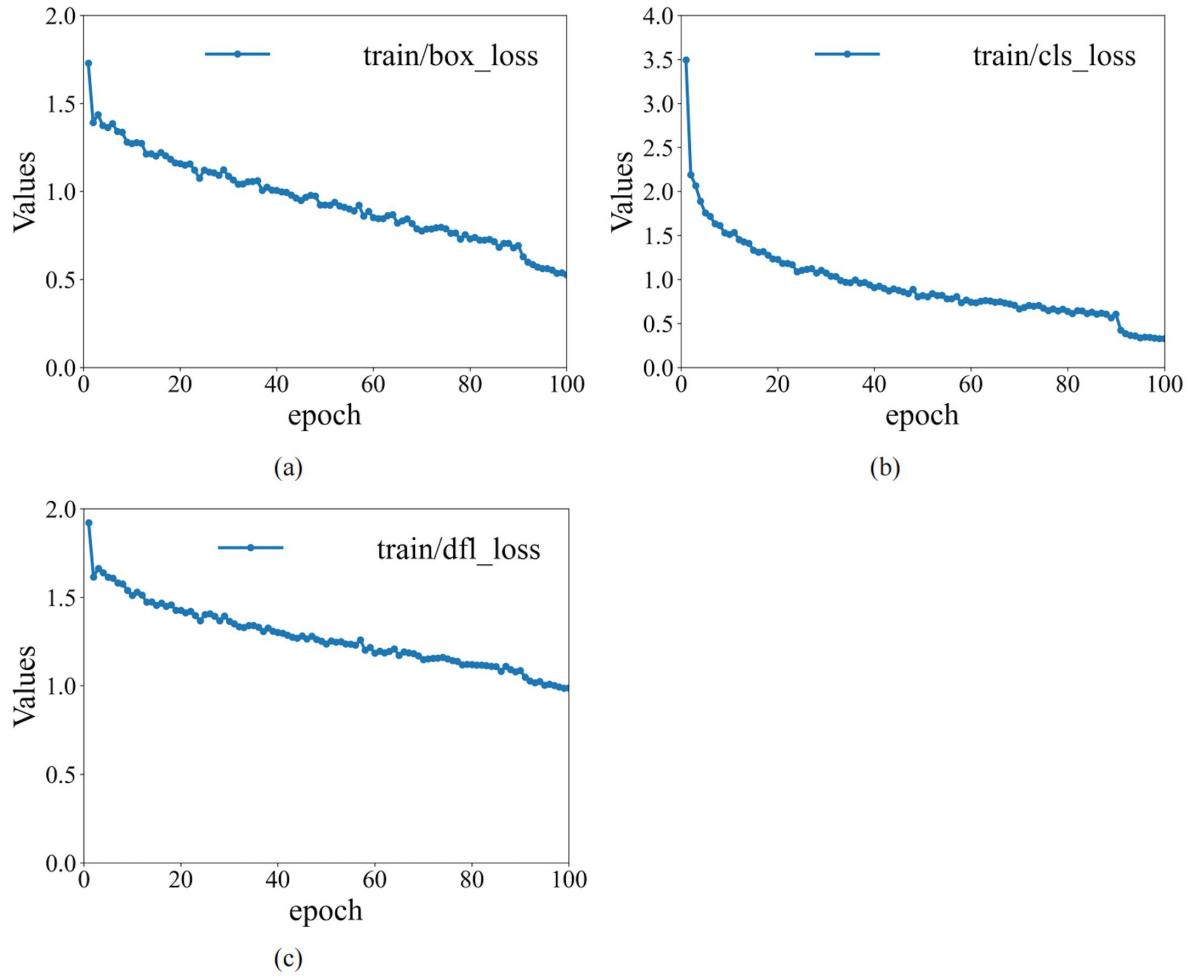


FIGURE 7. Loss of training data set

Figure 7 shows the changes in the loss functions of the training set. The horizontal axis represents the number of epochs, while the vertical axis corresponds to the loss values. Figure 7(a) shows the loss variation for bounding box predictions, which evaluates the discrepancy between the predicted and true bounding boxes. As observed in the figure, during the training process of bounding box anchoring, the overall loss convergence speed basically decreases linearly. This indicates that the YOLOv8 model progressively improves its ability to frame the Oracle-bone Inscription areas. The loss value gradually approaches the ideal value of 0 as the number of epochs increases, demonstrating that the model is learning to predict bounding boxes with increasing accuracy. Figure 7(b) shows the variation in classification loss for the training set, which measures the difference between the predicted and true categories. Overall, the loss value declines rapidly within the first 20 epochs, indicating that the model parameters are not yet fully optimized during the initial stages of training. However, this rapid decline indicates that the model quickly completes the low-precision classification tasks. As the number of epochs increases, the convergence rate of the classification loss gradually slows down. This suggests that the model has successfully captured the complex features of the five types of Oracle-bone Inscription data. After reaching 80 epochs, the loss stabilizes, indicating that the model has completed the classification task on the current dataset. Figure 7(c) shows the focal loss variation, an improved loss function designed to address challenges associated with imbalanced datasets and difficult-to-classify samples. Considering the small sample size selected in this study, this parameter was introduced to address class imbalance issues in object detection. The figure demonstrates that the overall loss converges at an approximately linear rate, reflecting an enhancement in the model's efficacy in handling hard-to-classify samples. The loss function design is intended to assist the model in focusing on Oracle-bone Inscription that are challenging to classify correctly. This approach is intended to enhance the model's classification ability for these challenging samples during the training process.

Figure 8 is a graph showing the changes in the loss function of the Oracle-bone Inscriptions validation set. The horizontal axis in the figure represents the number of epochs, while the vertical axis represents the corresponding loss value. Figure 8(a) shows the loss of the model when locating the bounding boxof the Oracle-bone Inscriptions. The loss decreases rapidly in the early stages of training and stabilizes thereafter, suggesting that the model efficiently learns the spatial distribution characteristics of the glyphs. By the later stages, the model achieves high positioning accuracy, reflecting its proficiency in bounding box localization. Figure 8(b) represents the loss for the classification task of Oracle-bone Inscriptions glyphs. The decreasing trend is gentler compared to Figure 8(a), and certain fluctuations occur during training. This could be attributed to the diversity and complexity of Oracle-bone Inscriptions, which make classification more challenging. As a result, the model requires more time to learn the distinguishing features of the different glyphs. Figure 8(c) represents the loss of the model when dealing with the problem of uneven distribution of Oracle-bone Inscriptions. After an initial significant decrease, the loss fluctuates around a low value, indicating that the model has made some progress in handling imbalanced data but still faces certain challenges. These phenomena comprehensively illustrate the learning dynamics of the Oracle-bone Inscriptions recognition model across different tasks during training. The model demonstrates rapid loss reduction in the early stages, followed by stable fluctuations in later stages as it converges. It reveals that the model gradually converges and adapts to the complexity of Oracle-bone Inscriptions recognition.

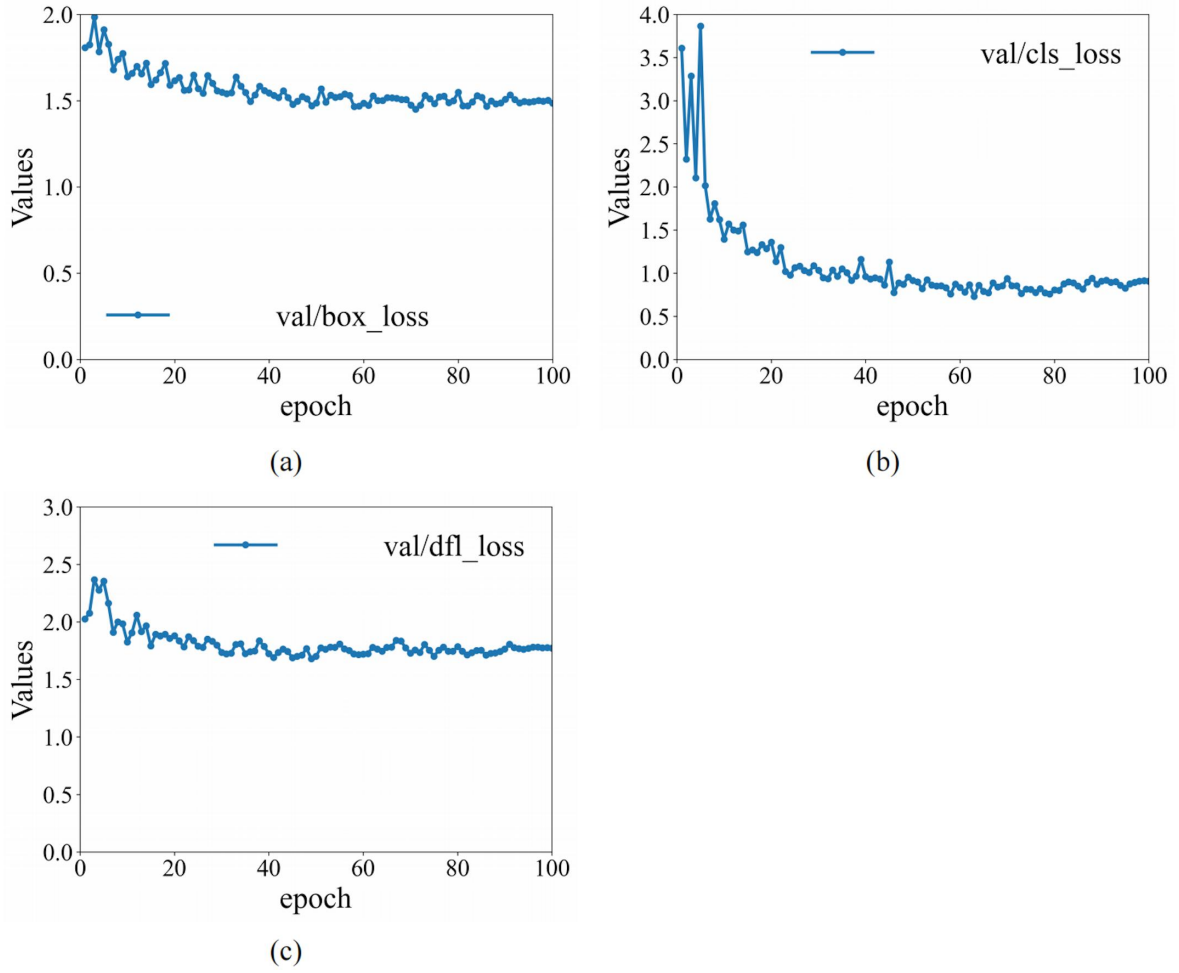


FIGURE 8. Loss of validation data set

Figure 9 presents four key performance metrics: mean Average Precision at IoU 50% (mAP50), mean Average Precision at IoU 50%-95% (mAP50-95), Precision, and Recall. These metrics reflect the model's learning dynamics and performance optimization process over the course of the training epochs.As shown in Figure 9(a), mAP50 increases rapidly in the early stages of training and stabilizes around epoch 20, approaching 1.0. This indicates that the model effectively captures the features of the Oracle-bone Inscriptions early on and achieve high recognition accuracy. As training progresses, the performance of the model gradually approaches saturation, with decreasing improvements, suggesting that the model is largely optimized at this stage. Figure 9(b) shows the trend of mAP 50-95. The value starts at 0 and gradually increases, but at a slower rate, and still does not reach 1.0 after 100 epochs. The mAP 50-95 calculation uses a stricter Intersection over Union (IoU) threshold, which demands higher precision in object recognition. Therefore, the model requires more time to optimize these stricter conditions and improve its performance under higher IoU requirements. Figure 9(c) shows the variation in Precision. Although there is significant fluctuation in Precision during the early stages of training, it gradually diminishes and stabilizes as training progresses. This fluctuation may be due to the model's unstable fit to the Oracle-bone Inscriptions data in the early stages. As training continues, the model gradually learns more accurate features, reduces false positives, and enhances its ability to recognize Oracle-bone Inscriptions. In Figure 9(d), Recall also shows significant fluctuations in the early stages of training, followed by gradual stabilization. The fluctuations in Recall may stem from the model's unstable

ability to identify positive samples of Oracle-bone Inscriptions in the early stages. As training progresses, the model continually improves its recognition of positive samples, thereby increasing Recall. In summary, the model's performance across different metrics shows varying optimization trends. This reflects its gradual adaptation to the features of the Oracle-bone Inscriptions data and its performance improvement through continuous optimization.

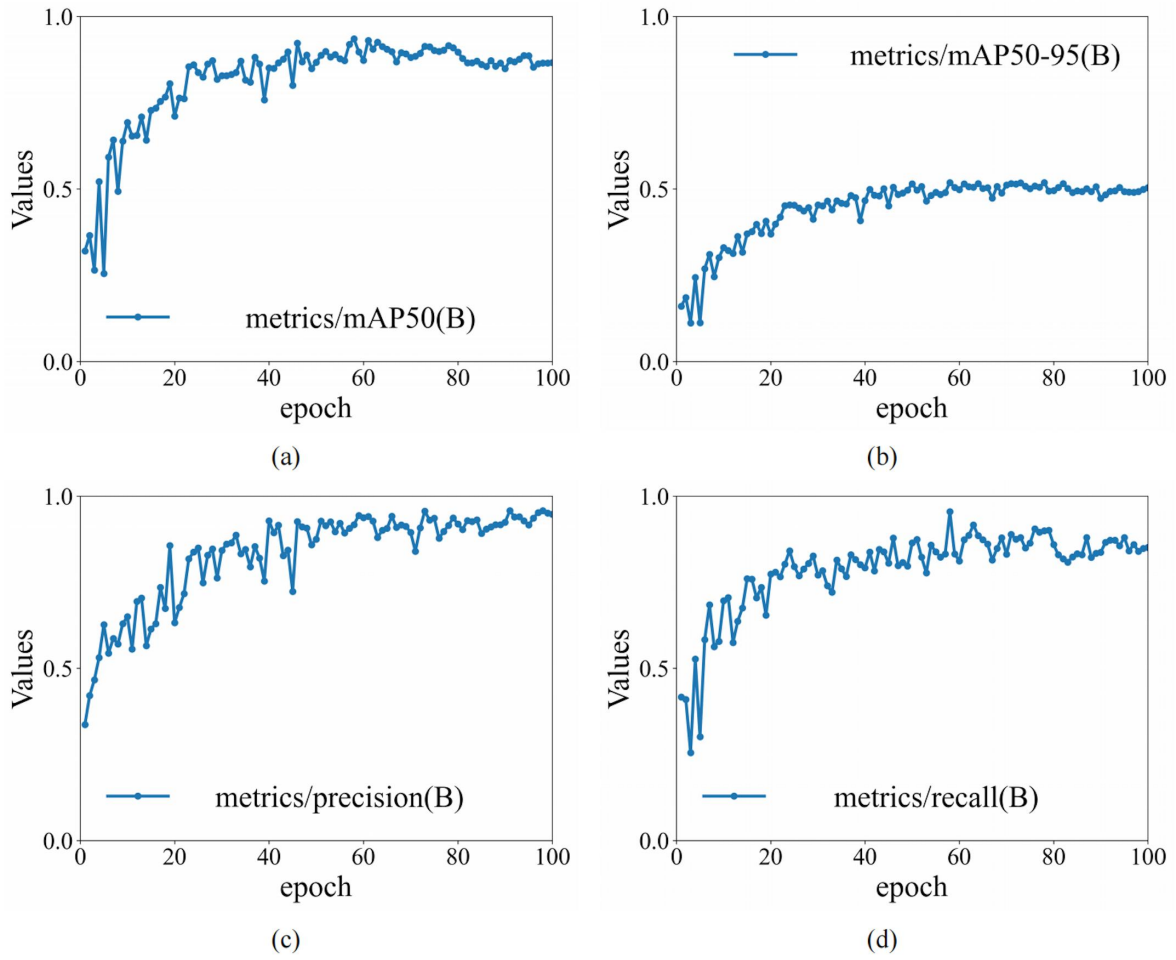


FIGURE 9. Average accuracy under different IoU thresholds

Figure 10 shows the variation in precision at different confidence levels. As the confidence threshold increases, precision generally increases, indicating that predictions with higher confidence are more reliable. Specifically, when the confidence threshold reaches 0.881, the average precision for all categories reaches 1.00. This means that at this confidence level, the model's predictions are almost completely accurate. There are differences in performance between categories. Some categories (such as 'bing' and 'zi') have higher accuracy at lower confidence levels. Other categories (such as 'yu' and 'geng') require higher confidence levels to achieve the same level of accuracy. This finding suggests that the performance of the model varies across different categories of data. This variation can be attributed to several factors, including the complexity of the category characteristics, the distribution of the sample, and the inherent discriminability of the data. At certain medium confidence levels, the detection precision is highest for the 'zi' characters, followed by 'jia' and 'bing'. In contrast, the detection precision for the 'yu' characters is the lowest in this confidence range, indicating that the model has some difficulty in recognizing this

category. Overall, the changes in precision at different confidence thresholds reveal the model's accuracy in recognizing Oracle-bone Inscriptions characters. This also suggests that the learning effectiveness of the model varies across different. In general, the changes in accuracy at different confidence thresholds reveal the accuracy characteristics of the model in oracle bone character recognition. This shows that the learning effect of the model is different for different categories.

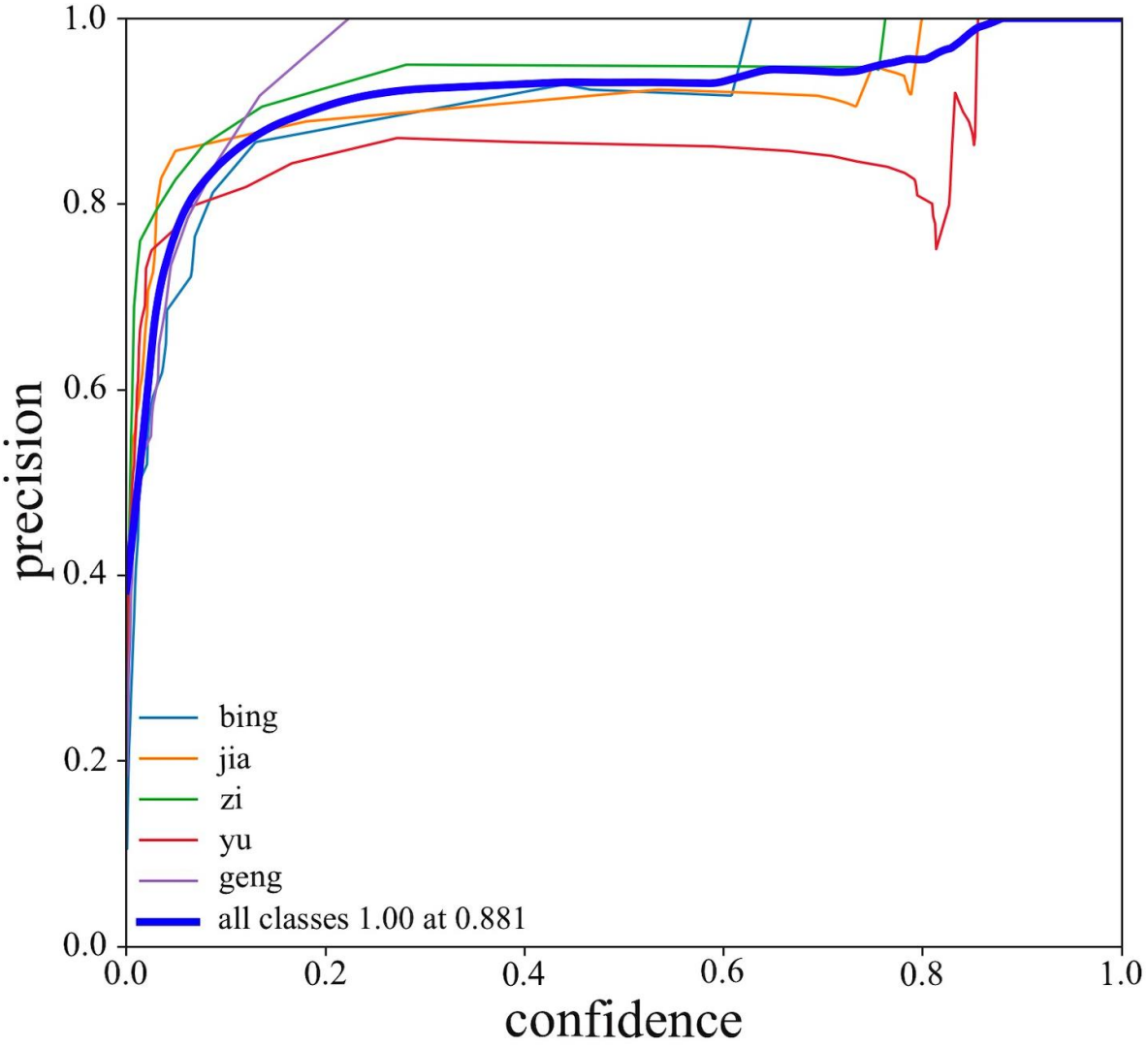


FIGURE 10. Precision-Confidence Curve

Figure 11 presents the recall of different categories ('bing', 'jia', 'zi', 'yu', and 'geng') varying with confidence,which are used to evaluate the performance of the object detection algorithm. When the confidence drops to 0, the recall for all target categories reaches 0.96, indicating that the model can identify most targets. As the confidence threshold increases, the recall for the 'bing' category remainshigh (around 0.96), indicating that the detection performance for this category is relatively stable. In contrast, the recall for the 'jia' and 'zi' categories decreases to around 0.75 and 0.65, respectively, which may be related to the similarity or complexity of their character features. In addition, the recall rates for the 'yu' and 'geng' categories drop sharply to around 0.2 and 0.3 respectively, and recognition ability is significantly reduced at higher confidence levels. Recognition was

significantly reduced at higher confidence levels, which may be related to the difficulty in distinguishing their visual features. Overall, there is a clear trade-off between recall and confidence. At lower confidence levels, the overall average recall is close to 1.0, indicating that the model can recognize most targets. At lower confidence thresholds, the overall average recall is close to 1.0, indicating that the model can detect most targets.

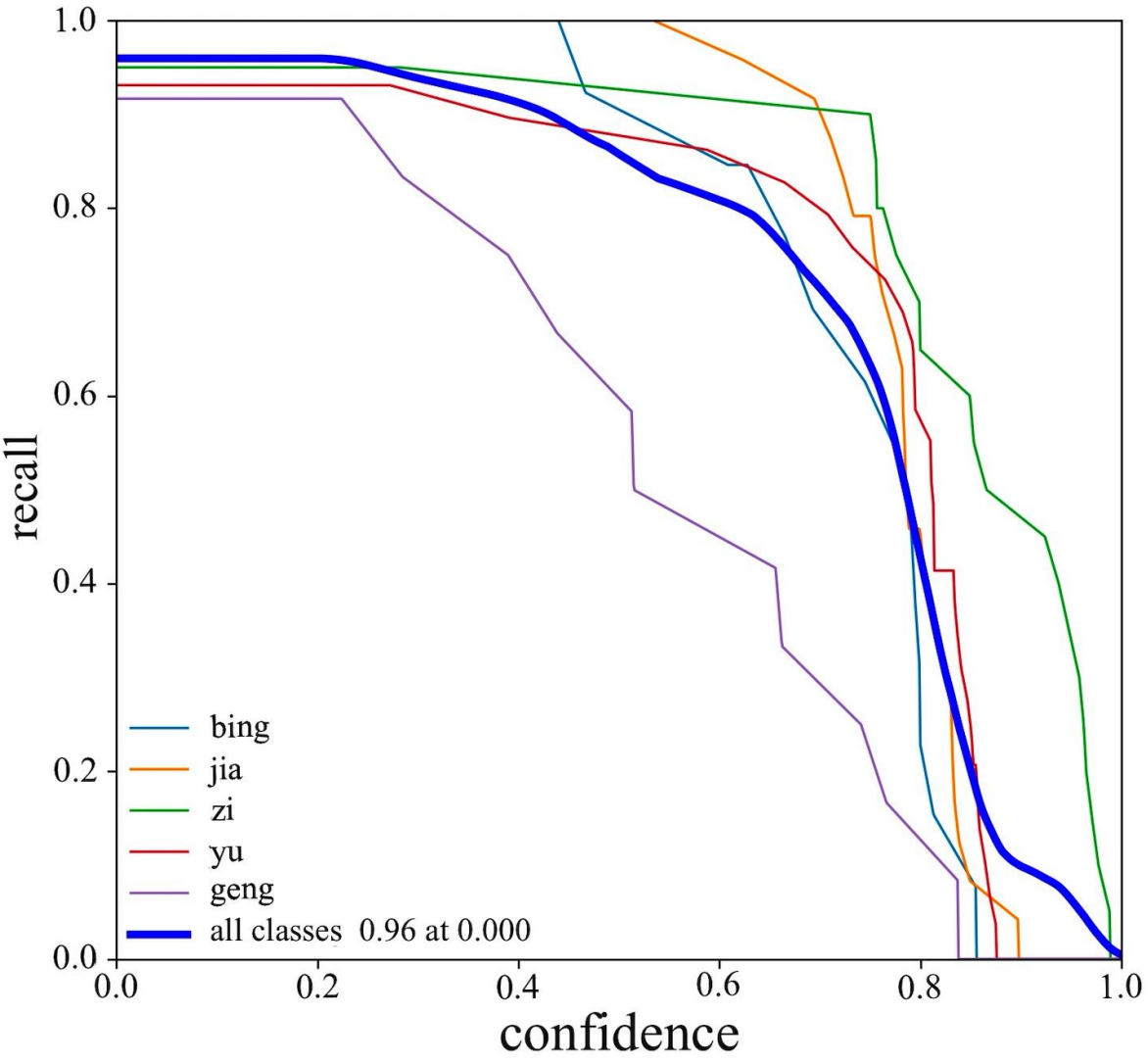


Figure 11. Caption Precision-Confidence Curve

Figure 12 shows the precision-recall curve, revealing the performance differences of different categories at different thresholds. The PR curve exhibits a gradually descending stepped trend, approaching the top-right corner. The overall mean average precision (mAP) is 0.935, indicating that the model performs well in Oracle-bone Inscriptions recognition. Specifically, the mAP for the 'jia' category is 0.963, indicating that this category maintains a high precision even at high recall rates. The mAP for the 'zi' category is 0.945, also showing good performance. The mAP for the 'geng' category is 0.923, slightly lower than 'jia' and 'zi', but still shows high precision and recall. However, the mAP for the 'yu' category is 0.860, indicating that precision drops at high recall rates. This suggests that the category may be associated with a higher number of false positive predictions. This may require adjusting the model parameters or augmenting the dataset to improve its precision.

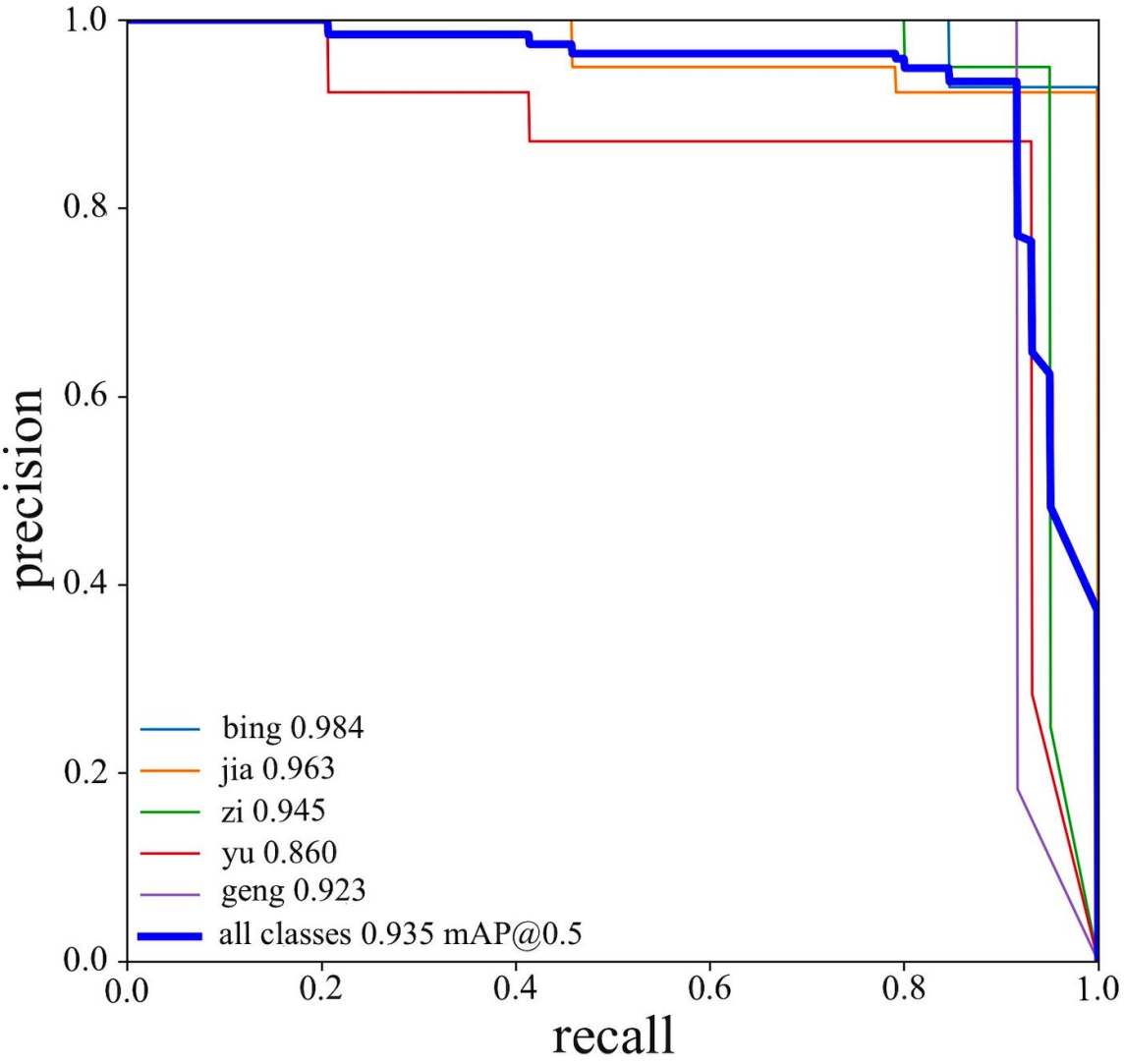


FIGURE 12. Precision-Recall Curve

Figure 13 presents a normalized confusion matrix diagram, which shows the performance evaluation junctions in different categories using the YOLOv8 detection algorithm. As shown in the matrix, the recall rate of the 'yu' category is 0.97, indicating that the model successfully detected 97% of the true 'yu' samples,thereby demonstrating a high degree of reliability. The correct classification rate for the background category is only 17% , suggesting that the model faces difficulties in classifying this category. This may be related to the unclear definition of background samples or the insufficient sample size in the data set. Furthermore, the recall rates for the 'bing' and 'geng' categories are 69% and 75% , respectively, which are relatively low. This suggests that these two categories may share feature similarities or that the training data is insufficient, making it difficult for the model to distinguish between them accurately. The off-diagonal elements reveal the confusion between categories. For example, 33% of the 'jia' category samples are incorrectly predicted as 'background'.

In the analysis, it was also noted that the last row of 'background' contains some misclassifications. Specifically, the values 0.31, 0.04, 0.10, 0.03, and 0.17 represent the probabilities that the algorithm incorrectly predicts true background samples as other categories. The values in the last column (0.13, 0.33, 0.13, 0.27, and 0.13) represent

the probabilities that the algorithm incorrectly predicts samples from other categories as background. The curves for 'bing' and 'jia' reach their highest point at a confidence threshold of approximately 0.236. This indicates that the optimal operating point differs for each category, and the fluctuations in the curves reflect the model's stability in detecting each category.

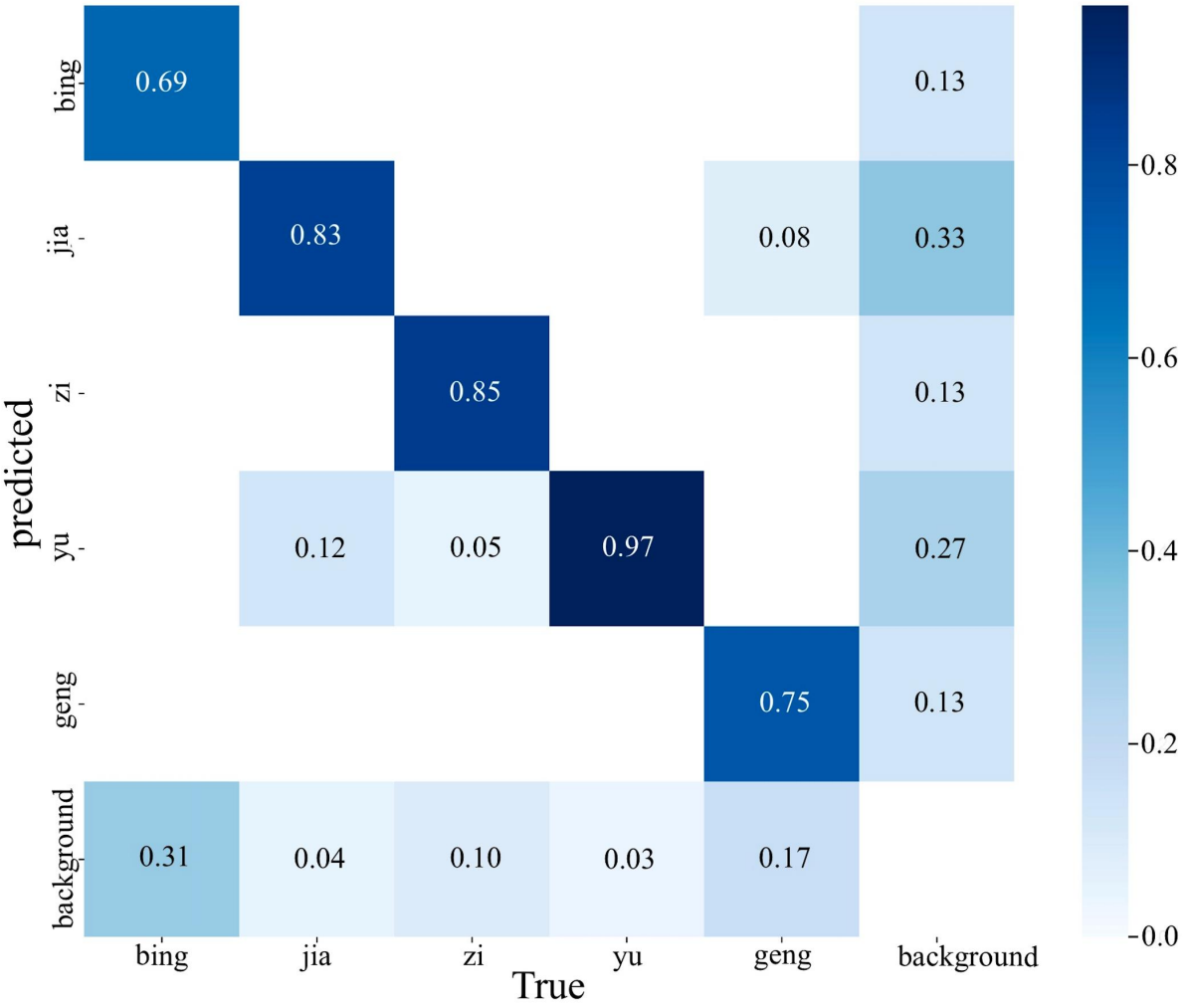


FIGURE 13. Normalized confusion matrix

The F1 score is the harmonic mean of precision and recall, used to evaluate the performance of a binary or multi-class system on imbalanced datasets. Figure 14 shows the variation of F1 scores for multiple categories at different confidence thresholds. The five curves in the figure represent the categories 'bing', 'jia', 'zi', 'yu', and 'geng', and as the confidence threshold increases, the F1 scores of each category exhibit different trends. When the confidence threshold is 0.93, the overall F1 score of the model is 0.236. The F1 curves for different categories reach their peak at different confidence thresholds.

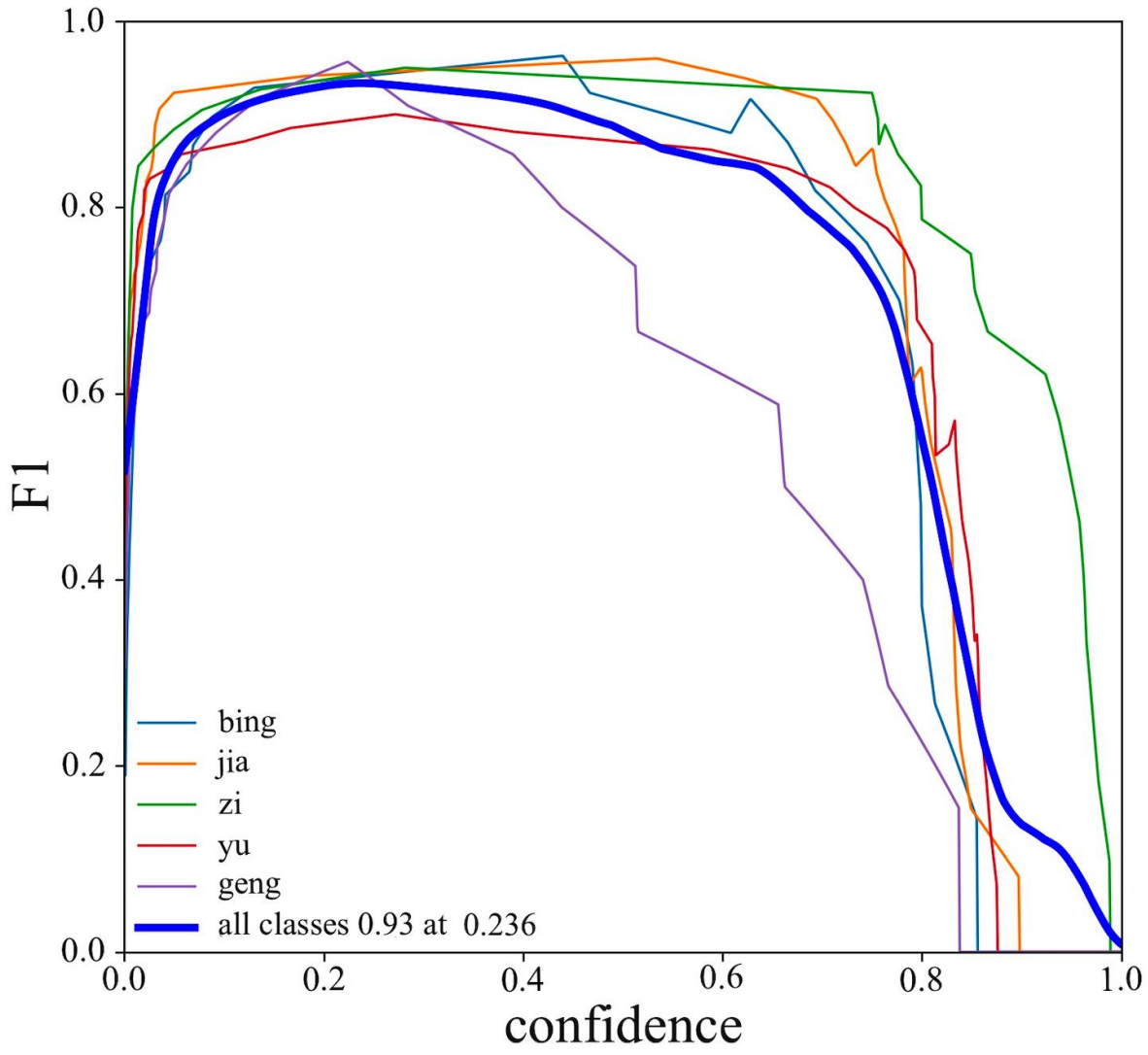


FIGURE 14. F1-Confidence Curve

5. **Experimental comparisons.**

In the same experimental environment, three distinct versions of the algorithmic models YOLOv8n, YOLOv8s, and YOLOv8x in YOLOv8 series are selected for a comparative analysis(including precision rate, recall rate, mean average precision and model size). The experimental results of performance comparison of different versions of YOLOv8 model are shown in Table 3 and Table 4.

(1) mAP (mean Average Precision): The mAP of YOLOv8s is 93.5%, which is higher than that of YOLOv8x (90.3%) and YOLOv8n (91.3%). The high mean average precision (mAP) of the YOLOv8s model indicates its strong capability in various target detection tasks.

(2) P (Precision): Although the YOLOv8s exhibits slightly diminished accuracy in comparison to the YOLOV8x, it nevertheless demonstrates superior performance (88.81%) in comparison to the YOLOv8n (86.1%).

(3) R (Recall): The YOLOv8s achieved a recall rate of 96%, which is the same as that of the YOLOv8x, but lower than YOLOv8n's 98%. The YOLOv8s shows a lower recall compared to the YOLOv8n, but it performs better in accuracy, which enhances its reliability in practical applications.

(4) Model size: The model size of YOLOv8s is 21.5 MB, which is 130.5MB of YOLOv8x and 6.2MB of YOLOv8n. The index 'model size' is used to describe the complexity and resource consumption of a given model. Smaller model sizes are generally more suitable for deployment in resource-constrained environments, such as mobile devices and embedded systems. The YOLOv8s model exhibits high performance while maintaining a relatively small model size, which renders it particularly well-suited for deployment in real-world applications.

To sum up, the YOLOv8s model is superior to the other two models in mAP and precision, and is equal to YOLOv8x in recall, and has a moderate model size, which makes it an excellent choice intarget detection tasks.

Table 3. Comparison of results from different models

Models	P/%	R/%	mAP/%	Algorithm volume size (106)
YOLOV8x	90.1%	96%	90.3%	130.5
YOLOv8n	86.1%	98%	91.3%	6.2
YOLOV8s	88.81%	96%	93.5%	21.5

The YOLOv8s model exhibits significant advantages over other models in the YOLOv8 series. In contrast to YOLOv8x, YOLOv8s has a significantly lower parameter count, decreasing from 68,157,423 to 11,137,535. This reduction in model complexity enables more efficient deployment on resource-constrained devices. In terms of computational complexity, the 28.7 GFLOPs of YOLOv8s is much lower than YOLOv8x's 258.1 GFLOPs, which reduces computational resource requirements. Although the inference time is slightly longer than YOLOv8x, the performance of 3.75ms is better than YOLOv8n's 4.06 ms, showing an optimal balance between performance and efficiency. In terms of GPU memory utilization, the YOLOv8s 2.59 GB is between YOLOv8x and YOLOv8n, providing a moderate memory occupation. Generally speaking, YOLOv8s demonstrates a significant advantage in resource constrained environments due to its lightweight and efficient design.

Table 4. Comparison between algorithms

algorithm	Layers	Parameters	GLOPs	Inference time/ms	GPU Memory/GB
YOLOv8x	365	68,157,423	258.1	1.05	8.41
YOLOv8n	225	3,011,823	8.2	4.06	1.47
YOLOv8s	225	11,137,535	28.7	7.75	2.59

6. **Conclusion.**

This study focuses on the recognition of Oracle-bone Inscriptions and uses three different YOLOv8 algorithm models to conduct a series of deep learning models experiments on Oracle-bone Inscriptions. The main conclusions are as follows:

(1) Data enhancement and optimization of the YOLOv8 algorithm

In the experiment of Oracle-bone Inscriptions recognition, t the YOLOv8 algorithm is adopted, and the recognition accuracy and robustness of the model are significantly.improved by data enhancement techniques. YOLOv8s performed well in both F1 score and mAP values, reaching 93% and 93.5%, respectively.

(2) Comparative analysis of YOLOv8 models In terms of mean average precision (mAP), the YOLOv8s model leads the other models with a mAP value of 93.5%, which is 3.2

percentage points greater than that of the YOLOv8x and 2.2 percentage points higher than that of the YOLOv8n. For recall rate, the recall rate of YOLOv8s is the same as that of YOLOv8n, which is 96%, lower than that of YOLOv8n, which is 9%. In terms of precision, the YOLOv8s achieved an accuracy rate of 88.81%, which is lower than 90.1% of YOLOv8x and higher than 86.1% of the YOLOv8n.In terms of model size, the YOLOv8s has a size of 21.5 MB, which is much smaller than 130.5MB of YOLOv8x, but larger than 6.2MB of YOLOv8n. Therefore, YOLOv8s offers superior model compression and enhanced applicability, all while maintaining high detection accuracy.

REFERENCES

[1] X. Fu, Z. Yang, Z. Zeng, Y. Zhang, and Q. Zhou, "Improvement of oracle bone inscription recognition accuracy: A deep learning perspective," *ISPRS International Journal of Geo-Information*, vol. 11, no. 1, p. 45, 2022.

[2] J. Yuan, S. Chen, B. Mo, Y. Ma, W. Zheng, and C. Zhang, "R-gnn: recurrent graph neural networks for font classification of oracle bone inscriptions," *Heritage Science*, vol. 12, no. 1, p. 30, 2024.

[3] X. Lin, S. Chen, F. Zhao, and X. Qiu, "Radical-based extract and recognition networks for oracle character recognition," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 25, no. 3, pp. 219–235, 2022.

[4] J. Li, Q.-F. Wang, K. Huang, X. Yang, R. Zhang, and J. Y. Goulermas, "Towards better long-tailed oracle character recognition with adversarial data augmentation," *Pattern Recognition*, vol. 140, p. 109534, 2023.

[5] Z. Guo, Z. Zhou, B. Liu, L. Li, Q. Jiao, C. Huang, and J. Zhang, "An improved neural network model based on inception-v3 for oracle bone inscription character recognition," *Scientific Programming*, vol. 2022, no. 1, p. 7490363, 2022.

[6] M. Wang, W. Deng, and S. Su, "Oracle character recognition using unsupervised discriminative consistency network," *Pattern Recognition*, vol. 148, p. 110180, 2024.

[7] F. Gao, J. Zhang, Y. Liu, and Y. Han, "Image translation for oracle bone character interpretation," *Symmetry*, vol. 14, no. 4, p. 743, 2022.

[8] A. Guo, Z. Zhang, F. Gao, H. Du, X. Liu, and B. Li, "Applications of convolutional neural networks to extracting oracle bone inscriptions from three-dimensional models," *Symmetry*, vol. 15, no. 8, p. 1575, 2023.

[9] X. Yue, H. Li, Y. Fujikawa, and L. Meng, "Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition," *ACM Journal on Computing and Cultural Heritage*, vol. 15, no. 4, pp. 1–20, 2022.

[10] X. Shi and X. Shen, "Oracle recognition of oracle network based on ant colony algorithm," *Frontiers in Physics*, vol. 9, p. 768336, 2021.

[11] A. Hazra, P. Choudhary, S. Inunganbi, and M. Adhikari, "Bangla-meitei mayek scripts handwritten character recognition using convolutional neural network," *Applied Intelligence*, vol. 51, no. 4, pp. 2291–2311, 2021.

[12] K. Verma and R. K. Sharma, "Recognition of online handwritten gurmukhi characters based on zone and stroke identification," *Sādhanā*, vol. 42, pp. 701–712, 2017.

[13] A. Hamplová, A. Lyavdansky, T. Novák, O. Svojše, D. Franc, and A. Veselỳ, "Instance segmentation of characters recognized in palmyrene aramaic inscriptions." *CMES-Computer Modeling in Engineering & Sciences*, vol. 140, no. 3, 2024.

[14] J. Guo, C. Wang, E. Roman-Rangel, H. Chao, and Y. Rui, "Building hierarchical representations for oracle character and sketch recognition," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 104–118, 2015.

[15] W. Gao, S. Chen, C. Zhang, B. Mo, and X. Liu, "Obm-cnn: a new double-stream convolutional neural network for shield pattern segmentation in ancient oracle bones," *Applied Intelligence*, vol. 52, no. 11, pp. 12 241–12 257, 2022.

[16] S. T. Nabi, M. Kumar, and P. Singh, "A convolution deep architecture for gender classification of urdu handwritten characters," *Multimedia Tools and Applications*, vol. 83, no. 29, pp. 72 179–72 194, 2024.

[17] M. Kumar, M. Jindal, R. Sharma, and S. R. Jindal, "A novel framework for writer identification based on pre-segmented gurmukhi characters," *Sādhanā*, vol. 43, pp. 1–9, 2018.

[18] G. Pengcheng, W. Jiangqin, L. Yuan, X. Yang, and M. Tianjiao, "Fast chinese calligraphic character recognition with large-scale data," *Multimedia Tools and Applications*, vol. 74, pp. 7221–7238, 2015.
[19] F. Gao, X. Chen, B. Li, Y. Liu, R. Jiang, and Y. Han, "Linking unknown characters via oracle bone inscriptions retrieval," *Multimedia Systems*, vol. 30, no. 3, p. 125, 2024.
[20] Y.-J. Chiu, Y.-Y. Yuan, and S.-R. Jian, "Design of and research on the robot arm recovery grasping system based on machine vision," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 4, p. 102014, 2024.
[21] Y.-X. Gu, Y.-J. Chiu, and Y.-J. Shih, "Application of convolutional block attention module in marine biodiversity research using convolutional blocks: A deep learning method for swimming crab recognition," *Journal of Network Intelligence*, vol. 9, no. 4, pp. 2438–2458, 2024.
[22] K. Sun, M. Gao, and B. Li, "Research on two interpolation algorithms for anthropometric data completion–a case study on the data of elderly males in fujian, china," *Journal of Network Intelligence*, vol. 7, no. 4, pp. 1083–1101, 2022.
[23] K. Sun and M. Gao, "A method for restoring the anthropometric database of elderly women for design," *Journal of the Chinese Institute of Engineers*, vol. 47, no. 4, pp. 456–471, 2024.
[24] K.-K. Sun, J.-W. Huang, Y.-Y. Yuan, and M.-Y. Chen, "Classification and recognition of the nantong blue calico pattern based on deep learning," *Journal of Engineered Fibers and Fabrics*, vol. 19, p. 15589250241270618, 2024.