

Article

An Oracle Bone Inscriptions Detection Algorithm Based on Improved YOLOv8

Qianqian Zhen ^{1,2}, Liang Wu ^{1,2} and Guoying Liu ^{1,2,*}

¹ School of Software Engineering, Anyang Normal University, Anyang 455000, China; qian_zhen@aynu.edu.cn (Q.Z.); 02031@aynu.edu.cn (L.W.)

² Henan Province Oracle Bone Culture Intelligent Industry Engineering Research Center, Anyang 455000, China

* Correspondence: guoying.liu@aynu.edu.cn

Abstract: Ancient Chinese characters known as oracle bone inscriptions (OBIs) were inscribed on turtle shells and animal bones, and they boast a rich history dating back over 3600 years. The detection of OBIs is one of the most basic tasks in OBI research. The current research aimed to determine the precise location of OBIs with rubbing images. Given the low clarity, severe noise, and cracks in oracle bone inscriptions, the mainstream networks within the realm of deep learning possess low detection accuracy on the OBI detection dataset. To address this issue, this study analyzed the significant research progress in oracle bone script detection both domestically and internationally. Then, based on the YOLOv8 algorithm, according to the characteristics of OBI rubbing images, the algorithm was improved accordingly. The proposed algorithm added a small target detection head, modified the loss function, and embedded a CBAM. The results show that the improved model achieves an F-measure of 84.3%, surpassing the baseline model by approximately 1.8%.

Keywords: oracle bone inscriptions; object detection; deep learning; YOLOv8

1. Introduction

The earliest known mature writing system in China is represented by oracle bone inscriptions [1,2]. Chinese character development and construction were significantly influenced by the characters of OBIs, and as a longstanding literary form, OBIs contain a wealth of information that holds significant value for our understanding of global history, character evaluation, and many other aspects [3]. In-depth research on these objects enables us to continuously explore the treasures of ancient civilization and thus contribute to the history and promotion of Chinese traditional culture. In 2017, OBIs were successfully entered into the Memory of the World Register, demonstrating that the cultural value and historical significance of OBIs are recognized worldwide. With the comprehensive improvement of public cultural understanding in society, it is increasingly important to improve and enhance the dissemination and utilization of OBIs, allowing them to move out of the realm of study and into the public domain. This will revitalize their legacy in the current era and gradually draw attention in the academic community for new research directions. One of the fundamental research tasks regarding OBIs is detection. Rapidly detecting OBIs using information technology significantly contributes to the acceleration of research and development in OBI studies. Many OBIs remained entombed within ruins until they were unearthed in 1899 [4]. Given the frequent incompleteness and low clarity of oracle bones that have endured years of burial underground, the precise detection and identification of OBIs has consistently posed a challenging task.

OBI detection falls under the umbrella of object detection, which constitutes an important research area in computer vision. The primary objective of this domain is to autonomously pinpoint the location of target objects and classify them in videos or images. Currently, OBI detection technology may be broadly classified into two groups: deep learning-based OBI detection and classical methods-based OBI detection. In terms



Citation: Zhen, Q.; Wu, L.; Liu, G. An Oracle Bone Inscriptions Detection Algorithm Based on Improved YOLOv8. *Algorithms* **2024**, *17*, 174. <https://doi.org/10.3390/a17050174>

Academic Editor: Frank Werner

Received: 5 March 2024

Revised: 17 April 2024

Accepted: 20 April 2024

Published: 24 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of traditional methods, Xiaosong Shi [5] used a method based on connected components for oracle bone script detection, but this method is not very effective regarding complex backgrounds or noisy oracle bone rubbing images.

As computer technology continues to evolve, object detection algorithms rooted in deep learning have emerged as the prevailing trend. In recent years, many experts and scholars, both domestically and internationally, have dedicated significant effort toward researching OBI detection by using deep learning techniques. OBI detection algorithms rooted in deep learning can be primarily categorized into two types: two-stage and one-stage detection algorithms. Of these, YOLO [6] and SSD [7] fall under the umbrella of one-stage detection algorithms, and Faster R-CNN [8], RFBNet [9], etc., are two-stage detection algorithms. RefineDet [10] is an improvement based on the SSD algorithm and effectively blends the strengths of both types. Although two-stage detection algorithms have higher accuracy, their implementation is complex and computationally intensive, the inference speed is slow, and they perform poorly when detecting small and dense objects. Conversely, one-stage detection algorithms are fast, accurate, and widely applicable. In research on one-stage detection algorithms, the YOLO series is used extensively across a range of diverse scenarios because of its excellent performance; these algorithms are lightweight, have a low background false-alarm rate, show good performance in detecting small objects, have strong model generalization abilities, and have a fast detection speed. For instance, YOLO can be used to detect objects in pictures taken by unmanned aerial vehicles [11], for traffic sign recognition [12], for steel surface defect detection [13], etc. Lin Meng [14] at Ritsumeikan University in Japan enhanced the SSD algorithm to improve the detection accuracy of small-font oracle bone characters. Jici Xing [15] constructed an OBI detection dataset and analyzed the detection performance of representative general detection frameworks (comprising Faster R-CNN, SSD, RefineDet, RFBnet, and YOLOv3 [16]) applied to this dataset, thus improving the best-performing YOLOv3, achieving an F-measure of 80.1%. Guoying Liu [17] proposed a multi-scale Gaussian kernel-based oracle bone detector. Xinren Fu [18] proposed a pseudo-label-based oracle bone script detection architecture. While the algorithm demonstrated promising performance on the training set, its detection accuracy was suboptimal on the validation and test sets. While these algorithms have enhanced the performance of object detection to a certain degree, there is still room for further refinement in terms of detection accuracy. According to research results in the literature [15], YOLOv3 performs well among mainstream deep learning models.

YOLOv8 is the latest version (as of 1 January 2024) of the YOLO series, released by Ultralytics in 2023. YOLOv8 adopts a more efficient network structure and more advanced algorithm technology, enhancing the precision and rapidity of object detection even further. YOLOv8 includes multiple models at different scales, and YOLOv8_n stands out as the most compact and efficient model within this series, offering the quickest performance among its counterparts.

In conclusion, to further improve the precision of detection, the present study focused on the uneven target size and significant noise of OBI rubbing images as well as real-time detection. Using the YOLOv8_n model as a baseline, our primary advancements were as follows:

1. Given the concentration of oracle bone script detection data, some oracle bone characters in certain images are small. To enhance detection performance for small objects, a detection head with a scale of 160×160 was added;
2. Some of the images in the OBI detection dataset are of poor quality, and the WIoU (Wise-IoU) [19] loss function prioritizes ordinary-quality anchor boxes; therefore, the loss function was modified to WIoU;
3. The noise in the OBI rubbing images is quite severe. To focus the model on recognizing key points—the oracle bone characters—and to reduce focus on other information, the CBAM (convolutional block attention module) [20] was embedded in the network architecture.

Given what is currently known in the field of deep learning, no target detection algorithm combines a small object detection head, WIoU, and a CBAM with YOLOv8, as proposed in this paper, and neither is there such a technology for OBI detection.

2. Materials and Methods

2.1. Experiment Dataset

2.1.1. Dataset Introduction

This experiment's OBI detection dataset comes from the Oracle Bone Inscription Information Processing Laboratory of Anyang Normal University. The OBI detection dataset is a set of annotated data used to train and evaluate OBI detection algorithms. The dataset contains OBI rubbing images and related annotation information. The image set contains all the image data. There are 8895 training annotations and 411 verification annotations, all saved in JSON files.

The oracle bone inscriptions have become blurry after being buried underground for a long time, causing varying degrees of corrosion in the bones. The dataset was obtained by scanning the rubbing images, and some images in the dataset contain a large quantity of noise and cracks, as shown in Figure 1. During the training process, these low-quality images can interfere with the model's ability to learn about image features.

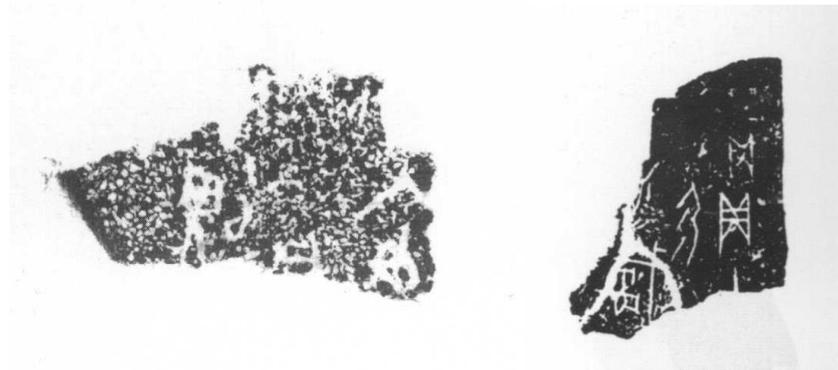


Figure 1. The noise and cracks.

The proportion of oracle bone annotation box sizes in the OBI detection dataset is shown in Figure 2. The sizes of the oracle bone characters vary, but they tend to be on the smaller side.

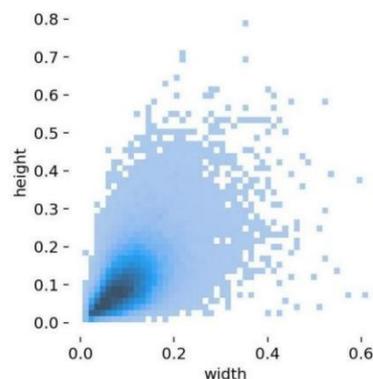


Figure 2. Aspect ratio of the bounding box size.

2.1.2. Processing Label Information

In the OBI detection dataset, the annotation file records the name of the image (`img_name`) and the specific position information of each OBI character. The annotation of the OBI character is implemented through a rectangular box, and the JSON file

contains the coordinates for the upper left and lower right corners of every bounding box, denoted as (x_0, y_0) and (x_1, y_1) , respectively. The bounding box in YOLOv8 saves the coordinates of the bounding box's center point, width, and height, denoted as $center_x$, $center_y$, $width$, and $height$, respectively. It is important to emphasize that the bounding box coordinate values have undergone normalization, resulting in a range between 0 and 1. At the same time, in the annotation set, the "height value" denotes the proportion of the annotation box's actual height in relation to the overall image height, whereas the "width value" signifies the ratio of the annotation box's actual width to the width of the image.

To ensure that YOLOv8 can accurately locate each oracle bone inscription, it is necessary to obtain the width and height parameters of the target image being processed, denoted as img_width and img_height , respectively. Then, the annotation information needs to be transformed according to Formulas (1)–(4).

$$center_x = \frac{x_0 + x_1}{2 \times img_width} \quad (1)$$

$$center_y = \frac{y_0 + y_1}{2 \times img_height} \quad (2)$$

$$width = \frac{x_1 - x_0}{img_width} \quad (3)$$

$$height = \frac{y_1 - y_0}{img_height} \quad (4)$$

Owing to the absence of categorical data pertaining to the characters within the dataset, this experiment treats all oracle bone characters as a single target and assigns them the same class label.

2.2. Input Data Assumptions

This study uses an improved YOLOv8 model to detect oracle bone inscriptions in the OBI detection dataset. To ensure effective training and accurate evaluation, the assumptions for the input data are as follows:

- **Image Format:**
All input images are in standard JPEG image format;
- **Color Space and Channels:**
The color space of the input images is RGB, and each image has three color channels;
- **Bounding Box Coordinate Representation:**
In the data preprocessing stage, all bounding box labels in the original JSON file are converted using Formulas (1)–(4);
- **Class Label Representation:**
The class label is represented by an integer, starting with 0. Since all oracle bone inscriptions in the dataset are treated as a single category in this study, it is assumed that 0 represents this unique class label;
- **Label Format and Transformation:**
The label file is converted into a txt file in the preprocessing stage; irrelevant label information from the original JSON file in the oracle bone inscription detection dataset is removed during data preprocessing, specifically parity bits; each oracle bone inscription's label information is on a separate line in the label file, consisting of the class label and the bounding box information; the data are separated by spaces;
- **Data Preprocessing:**
None of the input images have undergone any form of data preprocessing and are maintained in their unaltered state. The dataset has been cleansed to ensure a one-to-one

where (x_p, y_p) and (x_{gt}, y_{gt}) denote the central points of the prediction box and the ground-truth box, W_g, H_g are the width and height of the smallest enclosing box covering the two boxes, α is a positive trade-off parameter, and ν measures the consistency of the aspect ratio.

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w_p}{h_p} \right)^2 \quad (7)$$

The width and height of the ground-truth box are indicated by w_{gt} and h_{gt} , where w_p and h_p stand for the width and height of the prediction box, respectively.

$$\alpha = \frac{\nu}{L_{IoU} + \nu} \quad (8)$$

As a result of the unavoidable presence of substandard instances within the training data, the penalty for low-quality examples can be augmented by considering factors, ultimately leading to a decline in the model's generalization capabilities. In addition, when the overlap between the predicted box and the ground-truth box is very small or non-existent, CIoU may not provide effective gradient information. To tackle these challenges, this study uses the WIoU loss function. WIoU is a type of boundary loss based on the non-monotonic dynamic focusing method that evaluates the anchor boxes' quality through outlier evaluation and offers a prudent strategy for allocating gradient gains. The proposed strategy effectively diminishes the competitiveness of superior-quality anchor boxes and simultaneously mitigates adverse gradients emanating from inferior examples. As a result, WIoU concentrates on moderate-quality anchor boxes, leading to an improvement in the detector's overall performance. The distance attention is shown in Formulas (9) and (10).

$$L_{WIoU} = R_{WIoU} \cdot L_{IoU} \quad (9)$$

$$R_{WIoU} = \exp \left(\frac{(x_p - x_{gt})^2 + (y_p - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right) \quad (10)$$

To avoid generalizing the gradients with WIoU, which could potentially hinder convergence, the computational graph does not include W_g or H_g (denoted by the superscript *). This effectively removes the convergence-hindering factor. Therefore, there is no need to introduce new metrics. When the anchor boxes align well with the target box, WIoU diminishes the penalty imposed by the geometric factors, thereby allowing for minimal intervention during training and enhancing the model's generalization capability.

The primary step in implementing WIoU is computing the basic IoU value. Building on this, the final loss value is further refined by introducing distance attention mechanisms and dynamic adjustment factors. Subsequently, the obtained WIoU values are integrated into the loss function to replace the original CIoU loss component. During the model training phase, optimization adjustments are made based on the values of WIoU to more precisely calibrate the predicted bounding box positions and shapes.

2.4.3. Adding CBAM Attention Module

The complex background of OBI rubbing images and the small image area occupied by some of the detection targets lead to the presence of a significant proportion of extraneous features in the image iteration. To selectively enhance the feature channels that contain the most target information while also minimizing the disruption caused by invalid targets, enhancing the detection effect of the targets of interest, and thus enhancing the model's overall detection precision, this study embeds CBAM into the OBI detection algorithm. In recent years, attention modules have emerged rapidly, and common attention mechanisms include SE [24], GAM [25], EMA [26], and CBAM.

CBAM innovatively incorporates a spatial attention mechanism based on retaining the original channel attention mechanism, thereby comprehensively optimizing the network

in terms of both channel and spatial dimensions. This measure enables an optimized network to more accurately capture key features encompassing both channel-wise and spatial aspects, significantly improving the model’s efficiency in extracting features from both the channel and spatial dimensions. Figure 5 illustrates the structure of the CBAM.

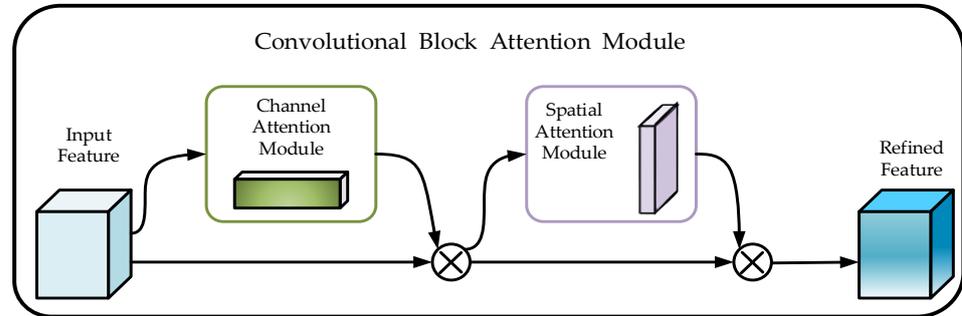


Figure 5. CBAM structure.

The CBAM processes the input feature map, F , by first obtaining its channel attention feature, $M_c(F)$. Multiplying the features element-wise with F yields the channel-refined feature F' . It then computes the spatial attention feature $M_s(F')$ of F' , multiplies it with F' element-wise, and obtains the final refined feature. This enhanced refined feature is used as the input for subsequent network layers, effectively preserving key information while suppressing noise and irrelevant information. Formulas (11) and (12) outline the entire attention procedure.

$$F' = M_c(F) \otimes F \tag{11}$$

$$F'' = M_s(F') \otimes F' \tag{12}$$

The input feature map undergoes max-pooling and average-pooling operations and feeds the resulting feature map into a shared multi-layer perceptron. Then, the feature maps generated by the shared MLP are combined, and the outcome is scaled by using the sigmoid function to derive the channel attention, M_c . Multiplying M_c with the feature map element-wise yields the channel-refined feature. The channel attention module’s architecture is schematically represented in Figure 6.

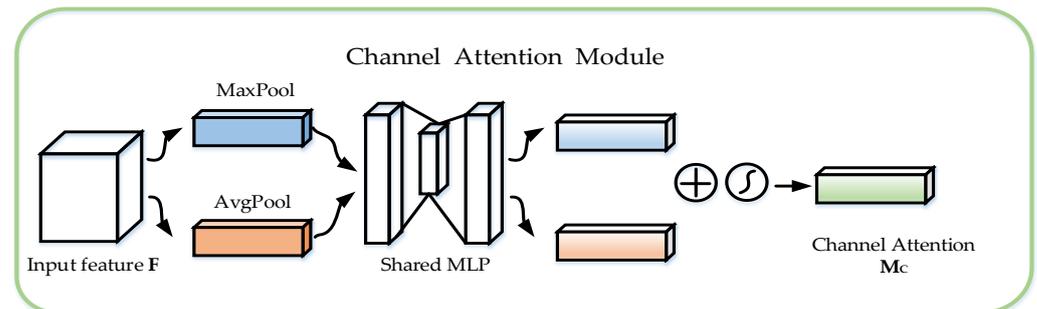


Figure 6. Channel attention module.

The channel attention module successively applies max-pooling and average-pooling to the refined channel features, stacks the results along the same dimension, adjusts the channel numbers using a convolutional layer, and utilizes the sigmoid function to derive spatial attention, M_s . Eventually, the channel-refined feature is element-wise multiplied with M_s to obtain the final refined feature. Figure 7 illustrates the spatial attention module’s structure.

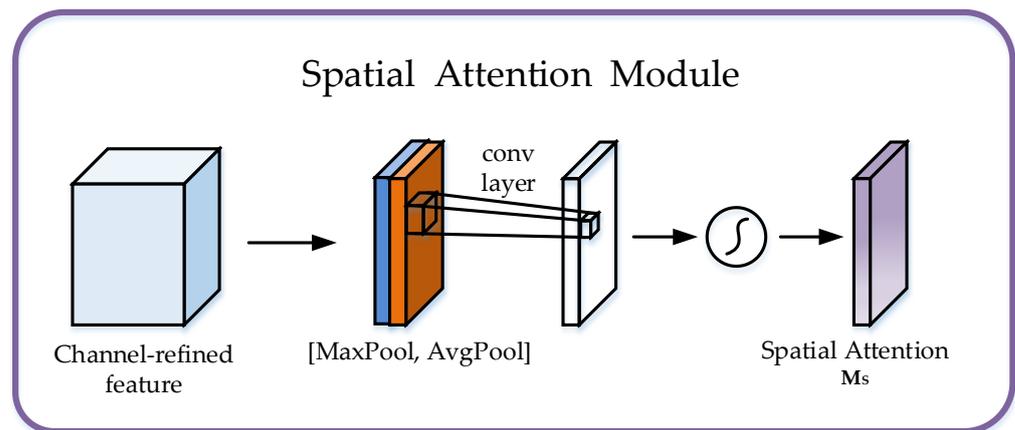


Figure 7. Spatial attention module.

CBAMs can be easily integrated into most mainstream networks, thus effectively improving the ability of network models to extract features without significantly increasing the computational complexity and number of parameters. Therefore, in this paper, the CBAM is embedded before the SPPF in the YOLOv8_n model.

2.5. Experimental Setup

2.5.1. Evaluation Metrics

Given that the dataset download page on the Yin Qi Wen Yuan website explicitly requires the use of F-measure as the metric for evaluating model detection performance, this study complies with this requirement. The experiment uses the balanced F-measure of precision and recall metric as the evaluation metric. Formulas (13)–(15) are calculated for precision, recall, and F-measure.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (14)$$

$$F - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

where TP denotes the count of accurately identified positive samples, FP represents the count of negative samples mistakenly labeled as positive, and FN signifies the count of positive samples incorrectly identified as negative.

2.5.2. Experimental Environment

The experimental environment utilized in this study is presented in Table 1.

Table 1. The experimental environment.

Item	Description
Operating System	Windows 11
Deep Learning Framework	PyTorch 1.11.0
CUDA Version	11.3.1
CPU Model	Intel(R) Xeon(R) Silver 4210R
GPU Model	NVIDIA GeForce RTX 3090
GPU Memory	24 GB
Programming Language	Python

2.5.3. Experimental Design

To thoroughly investigate the contribution of the model's improvements to its overall achievement, thus laying the foundation for future optimization work on the model, we designed an ablation experiment. In this experiment, a step-by-step introduction strategy was adopted to sequentially integrate the small target detection head, adjusted loss function, and an attention mechanism into the baseline model. Furthermore, to specifically study the effects of different attention mechanisms, GAM, SE, EMA, and CBAM were, respectively, embedded in the network's 9th layer on the basis of adding a small target detection head and adjusting the loss function.

Throughout the training phase, the OBI images in their original forms underwent scaling to achieve a resolution of 640×640 pixels. The learning rate commenced at 0.01 and remained unchanged, terminating at the same value. Every experimental iteration comprised 100 epochs, and the batch size for image processing was consistently maintained at 32. During training, all other parameters were set to the default values of the YOLOv8 model.

3. Results

3.1. Results of Ablation Experiments

Through carefully designed ablation experiments, this study systematically explored the impact of gradually incorporating small target detection heads, adjusting loss functions, and embedding attention mechanisms into a baseline model (YOLOv8_n) on overall performance. The experimental results for the validation set are shown in Table 2, which reveals the specific contributions of each component to the model's performance. Specifically, introducing small target detection heads enhanced the model's ability to recognize small targets to a certain extent, the WIoU significantly improved the model's detection performance, and introducing the CBAM further strengthened the model's ability to process key information. In summary, the model demonstrated a significant improvement in overall performance, providing a solid foundation for applications and subsequent research in related fields.

Table 2. The results of the ablation experiments.

Small Target Detection Head	WIoU	CBAM	Precision	Recall	F-Measure	FLOPs (GFLOPs)
			0.844	0.808	0.8256	8.2
✓			0.843	0.822	0.8324	12.6
	✓		0.873	0.811	0.8409	8.2
		✓	0.847	0.815	0.8307	8.2
✓	✓		0.847	0.825	0.8359	12.6
✓		✓	0.85	0.817	0.8332	12.7
	✓	✓	0.858	0.824	0.8407	8.2
✓	✓	✓	0.85	0.837	0.8434	12.7

To evaluate the usefulness of the attention module, GAM, SE, EMA, and CBAM were separately embedded into the 9th layer of the main network on the basis of the addition of a small object detection head and modification of the loss function. Given the experimental results presented in Table 3, it is evident that the F-measure value was highest when CBAM was embedded.

Table 3. The results of embedding the attention module.

Algorithm	Precision	Recall	F-Measure	FLOPs (GFLOPs)
YOLOv8_n + small target detection head + WIoU + GAM	0.866	0.821	0.8429	14.0
YOLOv8_n + small target detection head + WIoU + SE	0.856	0.829	0.8423	12.6
YOLOv8_n + small target detection head + WIoU + EMA	0.86	0.821	0.84	12.7
YOLOv8_n + small target detection head + WIoU + CBAM	0.85	0.837	0.8434	12.7

3.2. Result Analysis

3.2.1. Quantitative Analysis

- Adding a Small Target Detection Head Experiment

After integrating the small target detection head component, precision was almost unchanged compared with the baseline model, but the recall was enhanced by 1.4%, and the F-measure increased from 0.826 to 0.832. The improved recall rate indicates that the model's ability to identify true positives was enhanced, thereby enabling the more accurate reduction of false negatives. The experimental data show that the detection performance underwent a modest enhancement after adding the small target detection head. However, FLOPs increased by approximately 50% over the baseline model.

- Improving Loss Function Experiment

Changing the loss function from CIoU to WIoU resulted in a minor improvement in recall compared with the baseline model, with a 2.9% improvement in model accuracy. The improved accuracy indicates more precise model predictions, which enhance reliability, reduce misjudgments, and ultimately improve decision accuracy. The experimental results show a 1.53% increase in the comprehensive F-measure metric, showing that modifying the loss function is the most significant improvement in the entire algorithm. Some of the images in the OBI detection dataset are of poor quality, and the WIoU loss function prioritizes ordinary-quality anchor boxes, thus improving the overall detection performance.

- Adding Attention Module Experiment

After adding the CBAM to the ninth layer of the backbone network, the model's recall rate increased by 0.7%, the model's accuracy improved by 0.3%, and the comprehensive indicator F-measure increased by 0.5% compared with the baseline model. The images in the OBI detection dataset contain a lot of noise. The attention module enabled the model to concentrate on important information, so the detection performance was improved to varying degrees by incorporating different attention mechanisms. This improvement contributes to the model's ability to provide more comprehensive and reliable results.

- The Proposed Algorithm Experiment

In the ablation experiment—in addition to testing the contribution of three improvement strategies separately—the effects of any combination of two improvement strategies were also tested and compared with the improvement plan adopted in this study. The experimental outcomes indicate that the proposed improvement plan had the best effect, with a 1.78% increase in F-measure. However, the algorithm's time complexity also increased by approximately 50%.

To assess the generalization capability of the model, testing was conducted on the training and validation sets, as shown in Table 4. The improved model's F-measure was 1.76% higher on the training set and 1.78% higher on the validation set compared with YOLOv8. The results indicate that the improved model and YOLOv8 have the same generalization ability.

Table 4. Comparison between the training set and the validation set.

Algorithm	Precision		Recall		F-Measure	
	Train	Val	Train	Val	Train	Val
YOLOv8	0.882	0.844	0.868	0.808	0.8749	0.8256
The improved YOLOv8 (this study)	0.892	0.85	0.893	0.837	0.8925	0.8434

- Comparison of the Results of the Improved Algorithm with Other Algorithms

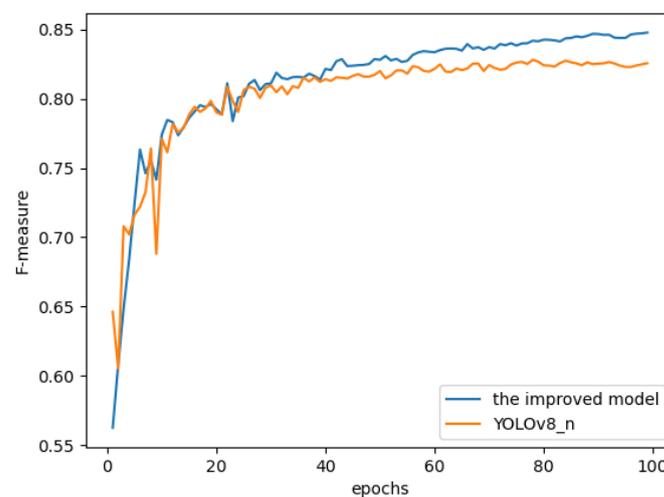
To further assess the model's performance, we compared the algorithm introduced in this study and various other object detection techniques. These algorithms include common first-stage and second-stage mainstream models as well as some efficient algorithms proposed for this dataset. The comparison results are presented in Table 5. The performance of the algorithm proposed in this study is much better than mainstream models like SSD and YOLOv3. Compared with the second-place Gaussian algorithm, the accuracy is 4.7% lower, but the recall rate is 6.2% higher. Therefore, the comprehensive F-measure index of precision and recall is 1.14% higher than the Gaussian algorithm. Compared with the latest U²NeT model, the F-measure on the training set decreased by 1.75%, but on the validation set, it increased by over 11%.

Table 5. Comparison between the improved algorithm and other existing algorithms.

Algorithm	Precision	Recall	F-Measure
YOLOv3	0.776	0.784	0.78
The improved YOLOv3 [15]	0.794	0.808	0.8009
Gaussian [17]	0.897	0.775	0.8315
SSD	0.748	0.758	0.753
Faster R-CNN	0.754	0.778	0.7658
RefineDet	0.752	0.805	0.7776
RFBnet	0.761	0.789	0.7747
U ² NeT [18]	0.738	0.721	0.7294
The improved YOLOv8 (this study)	0.85	0.837	0.8434

3.2.2. Qualitative Analysis

By sequentially introducing a small target detection head, improving the loss function, and integrating a CBAM, the improved algorithm was able to achieve a 1.78% improvement in F-measure compared with YOLOv8_n on the OBI dataset. The comparison results between the baseline and improved algorithms' F-measures are displayed in Figure 8.

**Figure 8.** Comparison of F-measure curves.

To comprehensively evaluate the performance of the model, further analysis of its precision–recall curves is needed. Figures 9 and 10, respectively, show the precision–recall curves of the baseline model and the improved model. The results indicate that the proposed improvement algorithm exhibits higher detection performance.

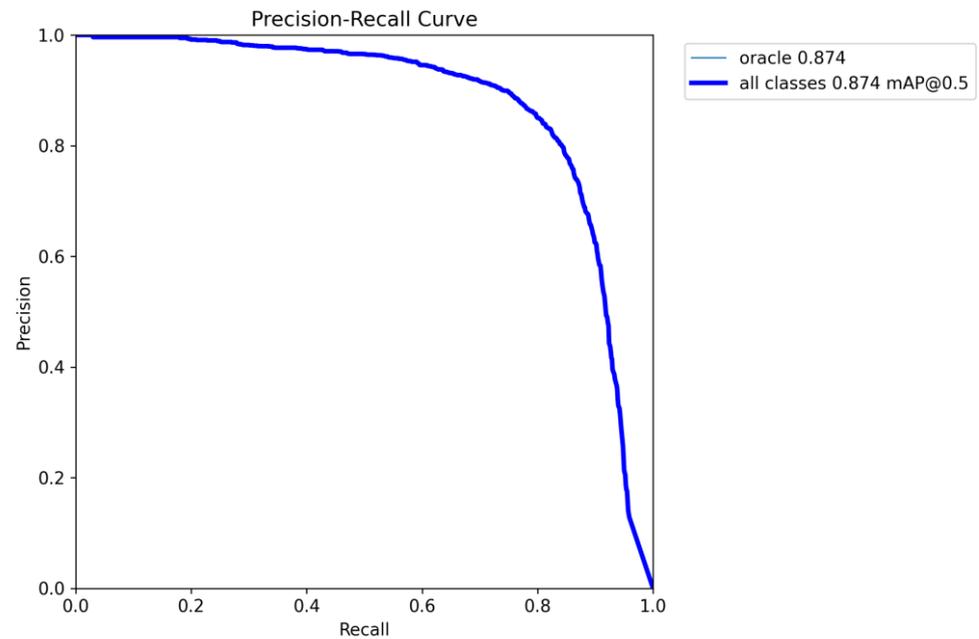


Figure 9. Precision–Recall Curve of YOLOv8_n.

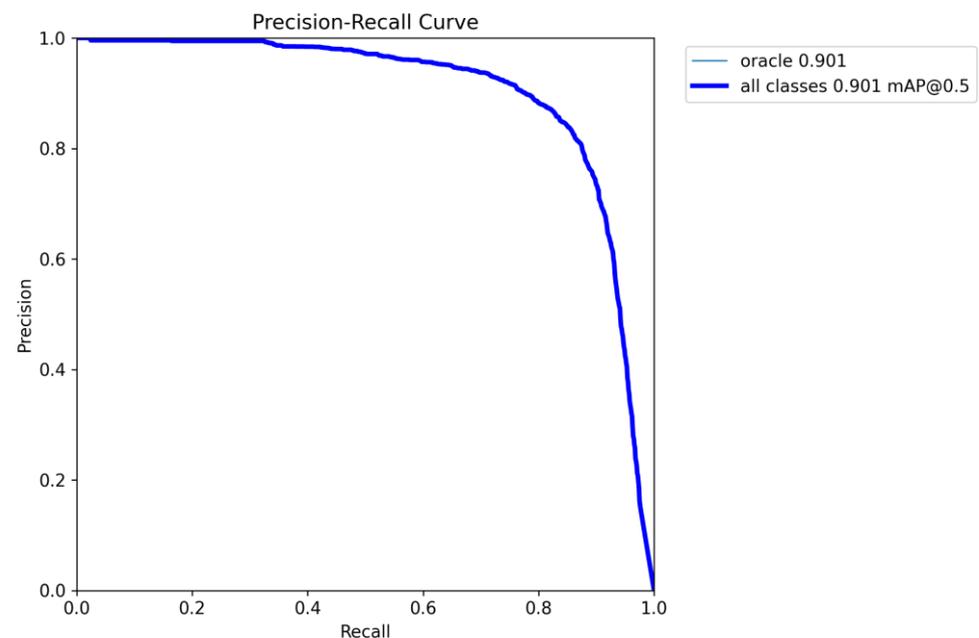


Figure 10. Precision–Recall Curve of the improved algorithm.

The annotated results were subjected to a rigorous comparison and in-depth analysis, the outcomes of which are depicted in Figure 11. The refined algorithm introduced within this research demonstrates enhanced detection accuracy when juxtaposed against the baseline model. For example, in the first oracle bone image, four oracle bone characters were manually marked, but the baseline model only detected one character, while the improved algorithm detected three characters. In the fourth image, six oracle bone characters were

marked manually; the baseline model detected four characters, while the improved algorithm detected all characters. Both the first image and the fourth image exhibited significant noise interference. Given the comparative analysis of the detection results, it can be seen that the improved algorithm can focus on the key detection region more effectively and therefore exhibits superior detection performance. The second and third images are not too noisy, and both the baseline model and the improved model detected all oracle bone characters, but the improved algorithm possessed a higher confidence level.

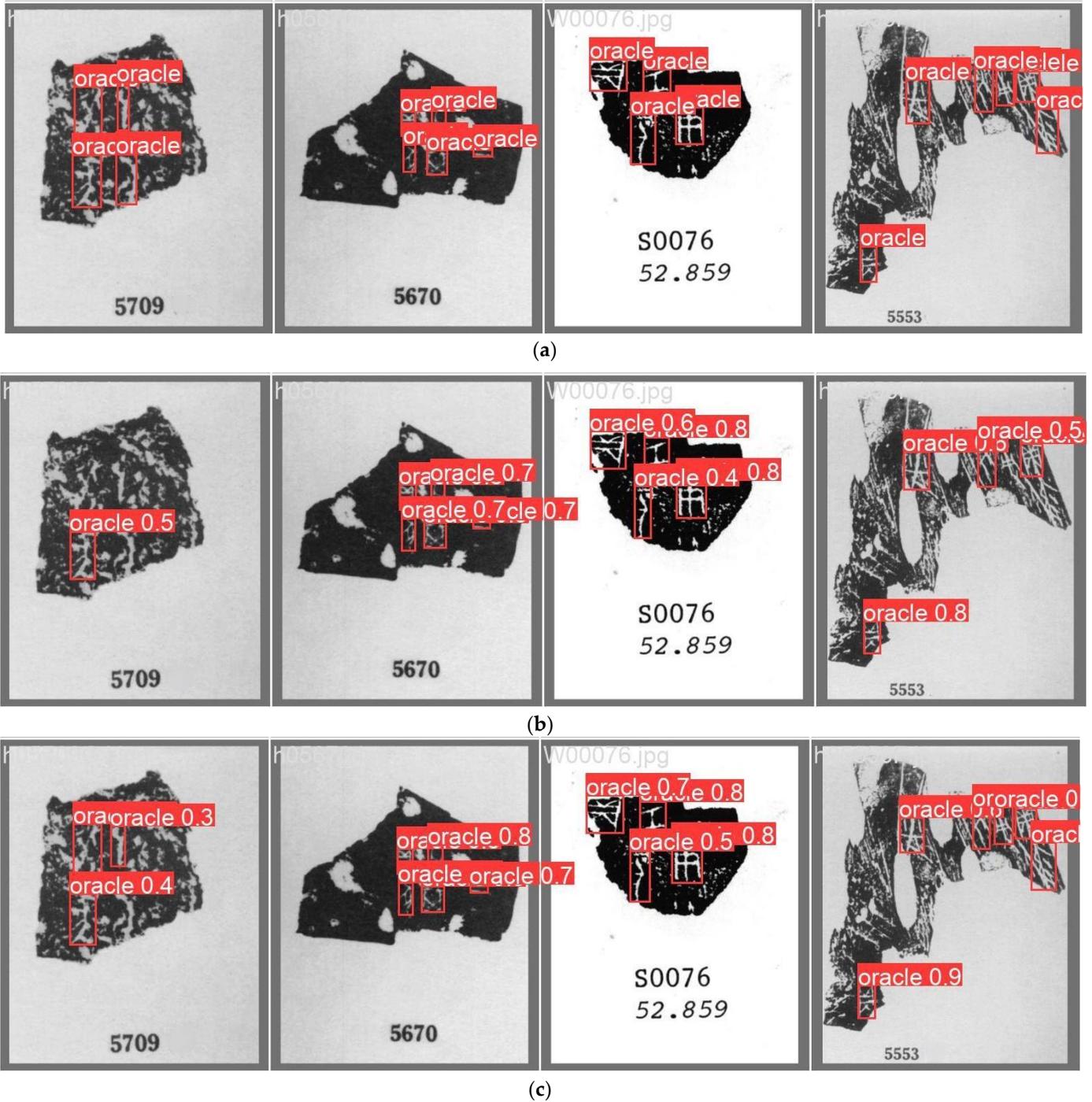


Figure 11. Comparison of detection results. (a) Manually labeled results. (b) Detection results for YOLOv8_n. (c) Detection results for the improved algorithm.

4. Discussion

Due to the large quantity of noise and cracks in the dataset, the improved algorithm's loss function was changed from IoU to WIoU and incorporated CBAM embedded into the backbone network. Compared with the traditional IoU loss function, WIoU provides a more accurate target detection assessment by more comprehensively considering the relationship between the target region and its surroundings. Its advantage is more pronounced when dealing with low-quality examples, as it effectively avoids overfitting the model to these examples, thereby enhancing the overall model performance. The CBAM consists of two key attention mechanisms—channel and spatial attention—allowing the network to dynamically adjust the feature extraction based on the characteristics of the current task and input image. Therefore, the proposed improved algorithm shows significantly better detection results for images with noise compared with the baseline model, leading to an overall improvement in detection performance.

The improved algorithm can meet the requirements of real-time detection in practical applications, but it also has certain limitations. Adding additional downsampling, C2f, and convolution operations to the model increases FLOPs decreases computational efficiency and prolongs model training time. The next research direction is finding ways to enhance detection performance for OBIs while ensuring computational efficiency.

To study the expandability of the proposed algorithm, an experiment was conducted on the NEU-DET open dataset. Given the lack of convergence after 100 epochs, the training rounds were increased to 200, keeping all other settings consistent with the OBI experiment. The experimental results are shown in Table 6. On the NEU-DET dataset, the detection performance of the improved algorithm increased by 3.13% compared with the baseline model. Therefore, the improved algorithm can be extended to detecting complex, small target background images in areas such as industrial defect detection and UAV aerial image analysis.

Table 6. Results for the NEU-DET dataset.

Algorithm	Precision	Recall	F-Measure
YOLOv8	0.773	0.684	0.7258
The improved YOLOv8 (this study)	0.807	0.713	0.7571

During the model training process, some technical challenges gradually emerged. The primary issue lies in the lack of strict one-to-one correspondence between the images and labels in the dataset, necessitating a comprehensive data-cleaning process to ensure a strict correspondence between images and labels. Additionally, the label format of the original dataset did not align with the requirements of the YOLOv8 algorithm. To address this challenge, it is necessary to accurately retrieve the actual dimensions of each image based on the label information of the dataset and recalculate the label data accordingly, thereby ensuring the effectiveness and accuracy of model training.

5. Conclusions

To enhance the efficacy of object detection algorithms for OBI images, we propose a refined version of the YOLOv8 algorithm. To boost detection accuracy for smaller objects, a dedicated small target detection head with a 160×160 scale was incorporated. To reduce the overwhelming influence of high-quality anchor boxes, minimize detrimental gradients arising from low-quality samples, and prioritize ordinary-quality anchor boxes, the loss function was improved to enhance the overall detection performance. To selectively enhance feature channels that contain the maximum target information, a CBAM was introduced. The improved YOLOv8 algorithm was verified on the OBI detection dataset, showing an approximately 1.8% increase in F-measure while maintaining a level of generalization ability comparable to the YOLOv8 algorithm. In actual testing, FLOPs increased

by approximately 50% compared with the baseline model. Our primary future work will be to further lighten the network model while also ensuring target detection performance.

Author Contributions: Conceptualization, Q.Z., L.W. and G.L.; methodology, Q.Z., L.W. and G.L.; software, Q.Z.; validation, Q.Z. and L.W.; formal analysis, Q.Z.; investigation, L.W.; resources, L.W.; data curation, Q.Z.; writing—original draft preparation, Q.Z.; writing—review and editing, Q.Z.; visualization, Q.Z.; supervision, L.W.; project administration, Q.Z.; funding acquisition, L.W. and G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Special Project for Cultural Research of Henan Xing Culture Engineering (No. 2022XWH115), the National Natural Science Foundation—Henan Province Joint Fund (No. U1804153), the general topic of the 2022 higher education scientific research plan of the China Association of Higher Education (No. 22cx0404), and the Anyang City Science and Technology Research Project (No. 2023C01GX019).

Data Availability Statement: The address of the dataset is <http://jgw.aynu.edu.cn>, accessed on 1 November 2023.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gao, J.; Liang, X. Distinguishing oracle variants based on the isomorphism and symmetry invariances of oracle-bone inscriptions. *IEEE Access* **2020**, *8*, 152258–152275. [CrossRef]
2. Jiao, Q.; Jin, Y.; Liu, Y.; Han, S.; Liu, G.; Wang, N.; Li, B.; Gao, F. Module structure detection of oracle characters with similar semantics. *Alex. Eng. J.* **2021**, *60*, 4819–4828. [CrossRef]
3. Fujikawa, Y.; Li, H.; Yue, X.; Aravinda, C.; Prabhu, G.A.; Meng, L. Recognition of oracle bone inscriptions by using two deep learning models. *Int. J. Digit. Humanit.* **2023**, *5*, 65–79. [CrossRef]
4. Yue, X.; Li, H.; Fujikawa, Y.; Meng, L. Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition. *ACM J. Comput. Cult. Herit.* **2022**, *15*, 1–20. [CrossRef]
5. Xiaosong, S.; Yongjie, H.; Yongge, L. Text on Oracle rubbing segmentation method based on connected domain. In Proceedings of the 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 3–5 October 2016; pp. 414–418.
6. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
7. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; Volume 28, pp. 91–99.
9. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
10. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-shot refinement neural network for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4203–4212.
11. Zhang, L.; Xiong, N.; Pan, X.; Yue, X.; Wu, P.; Guo, C. Improved object detection method utilizing yolov7-tiny for unmanned aerial vehicle photographic imagery. *Algorithms* **2023**, *16*, 520. [CrossRef]
12. Liu, T.; Dongye, C.; Jia, X. The Research on Traffic Sign Recognition Algorithm Based on Improved YOLOv5 Model. In Proceedings of the 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 6–8 January 2023; pp. 92–97.
13. Lu, Q.; You, C. Improved the Detection Algorithm of Steel Surface Defects Based on YOLOv7. In Proceedings of the 2023 4th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC), Nanjing, China, 18–20 August 2023; pp. 104–107.
14. Meng, L.; Lyu, B.; Zhang, Z.; Aravinda, C.; Kamitoku, N.; Yamazaki, K. Oracle bone inscription detector based on ssd. In Proceedings of the New Trends in Image Analysis and Processing—ICIAP 2019: ICIAP International Workshops, BioFor, PatReCH, e-BADLE, DeepRetail, and Industrial Session, Trento, Italy, 9–10 September 2019; pp. 126–136.
15. Xing, J.; Liu, G.; Xiong, J. Oracle bone inscription detection: A survey of oracle bone inscription detection based on deep learning algorithm. In Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya, China, 19–21 December 2019; pp. 1–8.
16. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
17. Liu, G.; Chen, S.; Xiong, J.; Jiao, Q. An oracle bone inscription detector based on multi-scale gaussian kernels. *Appl. Math.* **2021**, *12*, 224–239. [CrossRef]

18. Fu, X.; Zhou, R.; Yang, X.; Li, C. Detecting oracle bone inscriptions via pseudo-category labels. *Herit. Sci.* **2024**, *12*, 107. [[CrossRef](#)]
19. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
20. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
21. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
22. Du, S.; Zhang, B.; Zhang, P.; Xiang, P. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection. In Proceedings of the 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 16–18 July 2021; pp. 92–98.
23. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
25. Liu, Y.; Shao, Z.; Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv* **2021**, arXiv:2112.05561.
26. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. In Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–9 June 2023; pp. 1–5.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.