applied sciences

MDPI

Article

Gender Identification of Chinese Mitten Crab Juveniles Based on Improved Faster R-CNN

Hao Gu, Ming Chen * and Dongmei Gan

Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Hucheng Ring Road 999, Shanghai 201306, China; m220951602@st.shou.edu.cn (H.G.); m220901555@st.shou.edu.cn (D.G.)

* Correspondence: mchen@shou.edu.cn

Abstract: The identification of gender in Chinese mitten crab juveniles is a critical prerequisite for the automatic classification of these crab juveniles. Aiming at the problem that crab juveniles are of different sizes and relatively small, with unclear male and female characteristics and complex background environment, an algorithm C-SwinFaster for identifying the gender of Chinese mitten crab juveniles based on improved Faster R-CNN was proposed. This algorithm introduces Swin Transformer as the backbone network and an improved Path Aggregation Feature Pyramid Network (PAFPN) in the neck to obtain multi-scale high-level semantic feature maps, thereby improving the gender recognition accuracy of Chinese mitten crab male and female juveniles. Then, a self-attention mechanism is introduced into the region of interest pooling network (ROI Pooling) to enhance the model's attention to the classification features of male and female crab juveniles and reduce background interference on the detection results. Additionally, we introduce an improved non-maximum suppression algorithm, termed Softer-NMS. This algorithm refines the process of determining precise target candidate boxes by modulating the confidence level, thereby enhancing detection accuracy. Finally, the focal loss function is introduced to train the model, reducing the weight of simple samples during the training process, and allowing the model to focus more on samples that are difficult to distinguish. Experimental results demonstrate that the enhanced C-SwinFaster algorithm significantly improves the identification accuracy of male and female Chinese mitten crab juveniles. The mean average precision (mAP) of this algorithm reaches 98.45%, marking a 10.33 percentage point increase over the original model. This algorithm has a good effect on the gender recognition of Chinese mitten crab juveniles and can provide technical support for the automatic classification of Chinese mitten crab juveniles.

Keywords: Chinese mitten crab juveniles; target recognition; Faster R-CNN; Swin Transformer

check for updates

Citation: Gu, H.; Chen, M.; Gan, D. Gender Identification of Chinese Mitten Crab Juveniles Based on Improved Faster R-CNN. *Appl. Sci.* **2024**, *14*, 908. https://doi.org/10.3390/app14020908

Academic Editor: Douglas O'Shaughnessy

Received: 21 December 2023
Revised: 13 January 2024
Accepted: 15 January 2024
Published: 21 January 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/).

1. Introduction

The Chinese mitten crab [1], also known as the hairy crab or river crab, is native to East Asia. It predominantly inhabits rivers, estuaries, and coastal areas in China, Taiwan, South Korea, and Japan. Owing to its high market value and distinctive taste, it is regarded as an important commercial species and relished as a delicacy worldwide. China's river crab breeding output will reach 815,300 tons in 2022, with a total output value of more than 50 billion yuan (China Fishery Statistical Yearbook, 2022), playing an important role in its freshwater fisheries. Predominantly, pond culture, employing mixed-gender cultivation, is the principal method of Chinese river crab culture. Research shows notable differences in reproductive molting times and gonad development rates between males and females under these conditions [2]. Consequently, the market availability of adult male and female crabs varies. The optimal market period for female crabs in the Yangtze River Basin is from mid-to-late October to December. For male crabs, it is slightly later, extending from November to December [3]. Furthermore, due to differing reproductive molt times, male and female crabs experience molting at different times. Soft-shell crabs, usually those that have recently molted, are vulnerable to attacks from hard-shell crabs, leading to decreased survival rates. This presents a significant challenge in the single-sex cultivation of river crabs, necessitating the development of efficient sex identification technology [4]. Currently, manual identification of juvenile crabs is labor-intensive and prone to errors [5], highlighting the need for automated and efficient sex identification technology to enhance the accuracy and productivity of Chinese mitten crab culture and fishery management [6].

In recent years, the development of deep learning and computer vision technology has brought revolutionary changes to target detection [7]. Machine learning technologies are increasingly replacing manual labor, becoming a key focus in future production. Therefore, target detection algorithms based on deep learning emerged as the times required. This type of algorithm can generally be divided into two categories: two-stage target detection algorithms and single-stage target detection algorithms [8,9].

Two-stage algorithms, such as the R-CNN series (including R-CNN, Fast R-CNN, Faster R-CNN), Mask-RCNN, Cascade-RCNN, Libra-RCNN, and their variants [10–12], first generate candidate regions [13], then classify and regress bounding boxes in these regions. Known for their accuracy and robustness, they can precisely locate and classify targets through multiple iterations but are generally slower due to the additional candidate region generation step. On the other hand, single-stage algorithms like YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) perform classification and bounding box regression in one step. YOLO directly outputs object category and location, achieving end-to-end detection [14], while SSD predicts object location and category on different scale feature maps for enhanced accuracy and speed [15]. However, these single-stage algorithms, while faster, often lack the accuracy and robustness of two-stage algorithms and are prone to missed or false detections.

In this context, we have chosen to employ the R-CNN model. Our selection is driven by the model's proven efficacy in accurately detecting and classifying complex objects in diverse environments [16]. This choice aligns with our objective of differentiating between male and female Chinese mitten crab juveniles, a task that demands high precision due to the subtle physical differences between the sexes. The R-CNN model's robustness and accuracy, particularly in handling intricate image data, make it an ideal fit for our study. Furthermore, the model's ability to iteratively refine target location and classification enhances its suitability for our research aims.

As the scale of hairy crab farming becomes larger and larger, research on target detection and sex classification of live hairy crabs has increased in recent years [17]. In 2019, Zhao et al. used machine vision for the first time for river crab recognition [18]. They modified the inputs and outputs based on the deep convolutional neural network YOLO V3 used a self-constructed dataset for its training and realized high-precision recognition of underwater river crabs with an average precision mean value of 86.42%, an accuracy of 96.65% for the recognition of underwater river crabs, and a recall rate of 91.30%. In the following year, they again proposed a small live crab detector with lightweight Efficient Net as the backbone network, which requires only 15 MB of storage memory to achieve 96.21% accuracy and 94.86% recall [19]. To achieve faster and more accurate detection of pond culture river crabs, Ji W et al. proposed a crab target detection method based on the MobileCenterNet model [20], using the improved MobileNetv2 backbone network with a coordinate attention module for crab feature extraction, which both achieves lightweight and focuses the model's attention on crab-related features. The average precision and F1 value of MobileCenterNet are 97.86% and 97.94%, respectively, the model size is only 24.46 M, and the detection speed reaches 48.18 frames/s.

However, the current research on the sex of Chinese mitten crab is still not too much. Wei et al. proposed a multi-group convolutional neural network-based pike crab sex recognition method, which is based on the recognition of local images in the target image of the given strategy matching template by pattern matching [21], the navel of male pike crab is pointed and narrow, while the navel of the female is round and wide. Based on this difference, the final classification accuracy, recall, and checking accuracy of the method reached 95.59%, 94.41%, and 96.68%, respectively, and the recognition error rate was only 4.41%. Later, Cui Y et al. argued that male and female Chinese mitten crabs have different nutrients [22], so it is necessary to distinguish the sex of the crabs before they are sold in the market, and he proposed that we cannot rely only on the navel of the crab to distinguish the sex, so we combined the features of the crab abdomen and the shell to construct the gender classification system of the crabs using CNN. The classification accuracy of the model reached 98.90%. Although this method is more accurate for the identification of adult crabs, when the object of identification is changed to crab juveniles, its less obvious navel feature will seriously affect the identification results.

To address these problems, the primary solutions and contributions of this study are summarized as follows:

1. Proposing a model structure based on Faster R-CNN to enhance gender identification accuracy in Chinese mitten crab juveniles, especially in scenarios prone to misjudgments.
2. Integrating the backbone feature extraction network Swin Transformer with PAFPN to augment the model's feature extraction capability.
3. Introducing an attention mechanism in the ROI layer to further refine the accuracy of the target detection frame.
4. Employing an improved non-maximum suppression algorithm and focal loss function to train the model, thereby focusing more on samples that are challenging to distinguish.

The rest of this paper is organized as follows. Section 2 begins with a detailed description of the materials and methods used in the study, including data collection and image processing techniques. The second half of Section 2 elaborates on the proposed C-SwinFaster model, discussing its components and their functionality. Section 3 presents the experimental setup and results, demonstrating the performance of the model. Section 4 further demonstrates the effectiveness of the algorithm through comparison with other methods and ablation experiments. Finally, Section 5 summarizes the paper's findings, contributions, and potential areas for future research.

2. Materials and Methods

2.1. Materials

2.1.1. Image Acquisition and Annotation

Data set collection is crucial for employing deep learning models in target detection. Due to the scarcity of public data sets for the subject of this study, we undertook the task of data preparation. We collected images of artificially cultured Chinese mitten crab juveniles in the waters of Yangcheng Lake, Kunshan City, Jiangsu Province (31°25′34.716″ N, 120°53′10.752″ E). Examples of collection sites and image acquisition equipment are shown in Figure 1 [23]. To quickly obtain samples, crab juveniles were placed in batches into the pool during the collection process. The Canon PowerShot SX730 HS camera was placed in the pool using a stand, and the camera was photographed from a distance of 0.5 m, with the crab juveniles as the main focus, and saved in JPG format with a resolution of 1920 × 1440.

This article uses Labelme for gender classification area labeling [24]. During the marking process, individuals were manually identified by creating bounding boxes and coordinate axes, as demonstrated in Figure 2. The images of marked crabs are stored in JSON format. To meet our experimental requirements, we use a Python program to convert these marked images and labels into the COCO2017 format [25], resulting in the creation of instance_train2017.json and instance_val2017.json files.

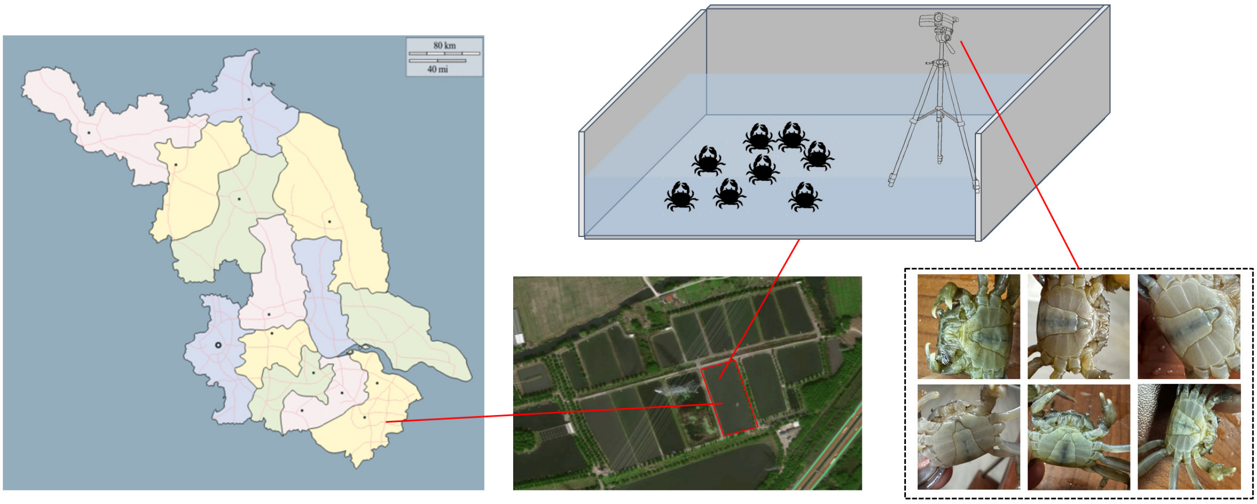

Figure 1. Acquisition site and examples of crab juveniles.

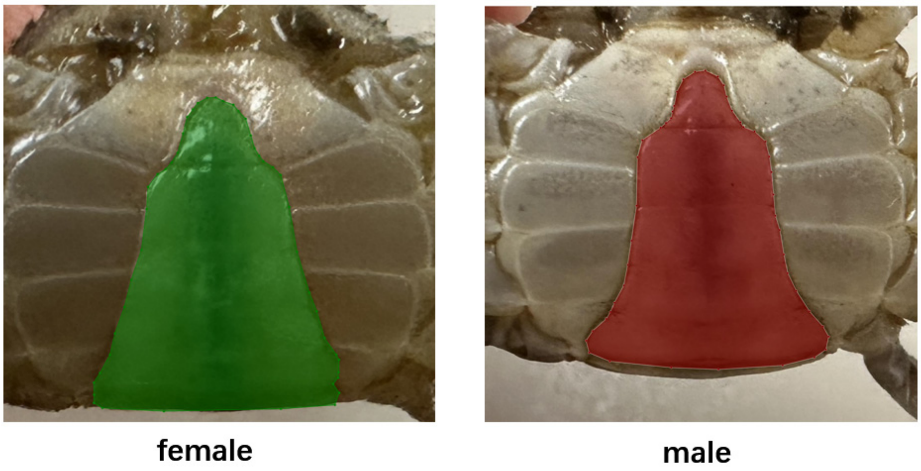

Figure 2. Annotation examples.

2.1.2. Image Augmentation

Deep convolutional neural networks excel in various computer vision tasks, including object detection. However, these networks usually require a large number of training images. Fine-tuning images is crucial to ensure that the loss function converges to the global minimum, enhancing the model's detection efficiency. However, obtaining large-scale data is often challenging in actual data collection. To overcome this and improve the model's detection capability and robustness, we augmented the existing dataset. Analysis of the crab juvenile images indicated that factors like shooting angle and lighting significantly influenced the data. Initially, data augmentation involved adjusting image brightness and adding noise, increasing the image count from 540 to 1620. Subsequent training revealed that while this first augmentation enriched target information, the 1620 images were insufficient, leading to potential model overfitting and reduced generalization capability. Therefore, we performed a second augmentation stage, including rotation, cropping, and flipping, resulting in 6480 images for training in this experiment. Table 1 compares the Chinese mitten crab juveniles dataset before and after enhancement.

Table 1. Comparison of crab juveniles data sets before data augmentation.

Data	Female	Male
Original data	280	260
First data augmentation	840	780
Second data augmentation	3360	3120

2.2. Overall Process Flow of the Proposed Method

This paper proposes a method for detecting Chinese mitten crab juveniles and classifying crab gender. The overall process flow is displayed in Figure 3.

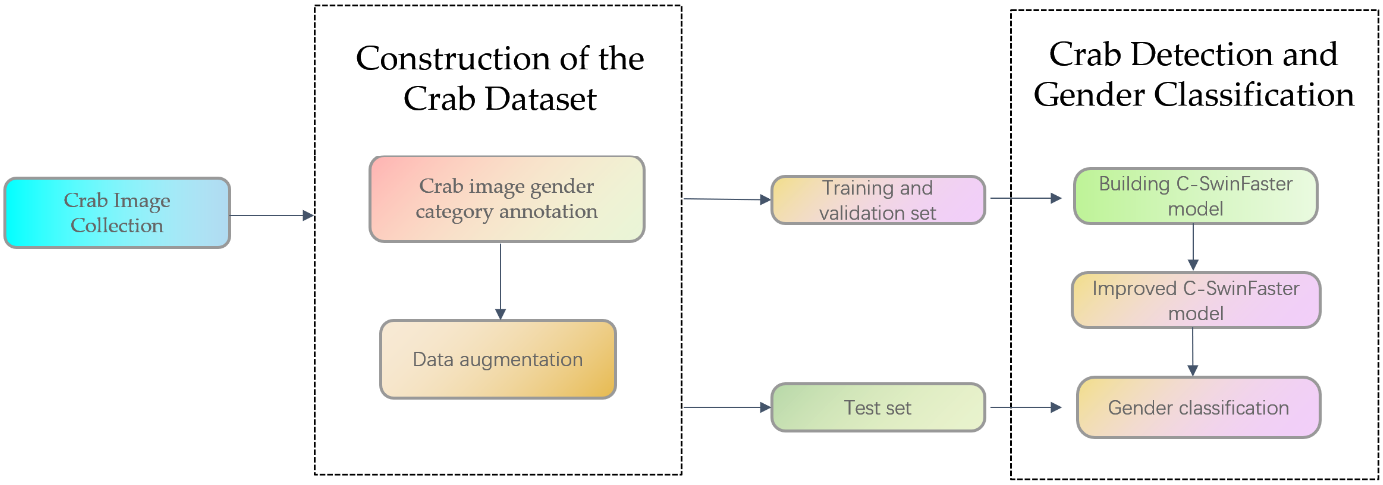
Crab Image Collection
Construction of the Crab Dataset
Crab image gender category annotation
Data augmentation
Training and validation set
Test set
Crab Detection and Gender Classification
Building C-SwinFaster model
Improved C-SwinFaster model
Gender classification

Figure 3. Overall process flow of the proposed method.

2.3. Methods

The C-SwinFaster model, proposed in this study, is an improvement based on Faster R-CNN. It includes a feature extraction network (Backbone), region recommendation network (RPN), region of interest pooling (ROI pooling), and classification regression module. First, we integrated the extraction network with an improved path aggregation feature pyramid network (PA-FPN), followed by processing through the RPN network and ROI pooling. Finally, the detection head carries out classification and regression tasks, enabling the effective identification of male and female mitten crabs. The model structure is shown in Figure 4.

2.3.1. Swin Transformer

The Swin Transformer is a potent and versatile deep-learning model, rooted in the Transformer network architecture. It adeptly captures global context information through a self-attention mechanism. It has demonstrated excellent performance and model versatility in tasks such as speech recognition, image classification, and object detection. Especially in the field of visual recognition, its superiority in handling targets with numerous small variations outshines traditional network structures [26]. Unlike models restricted by a fixed kernel size, it adopts a fully adaptive approach, efficiently dealing with sparse and irregularly shaped targets. For instance, in gender identification of Chinese mitten crab juveniles, it can precisely analyze and distinguish subtle edge details of different specimens.

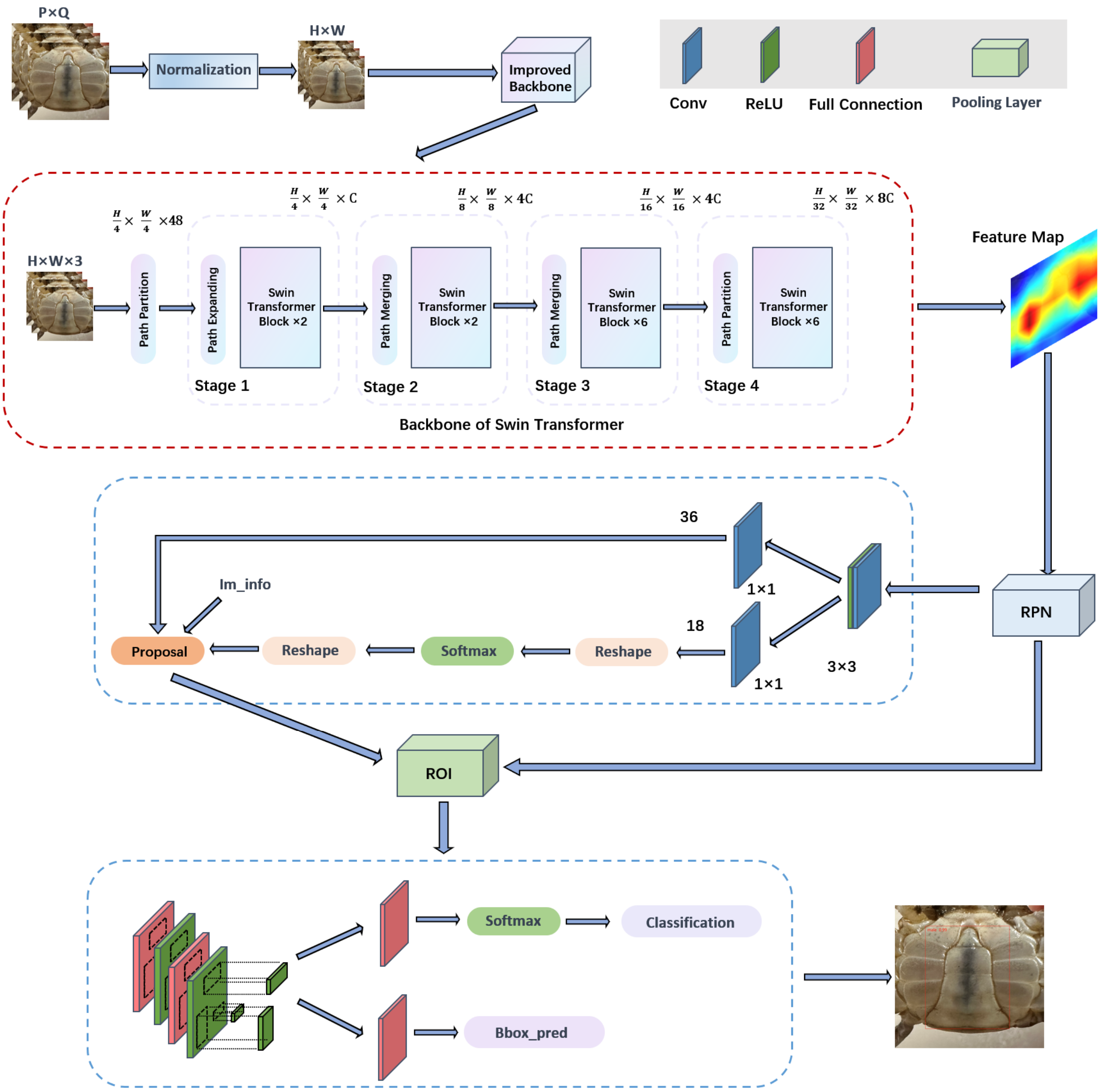
P×Q
Normalization
H×W
Improved Backbone
Conv
ReLU
Full Connection
Pooling Layer
H×W×3
$\frac{H}{4}\times\frac{W}{4}\times 48$
$\frac{H}{4}\times\frac{W}{4}\times C$
$\frac{H}{8}\times\frac{W}{8}\times 4C$
$\frac{H}{16}\times\frac{W}{16}\times 4C$
$\frac{H}{32}\times\frac{W}{32}\times 8C$
Path Partition
Path Expanding
Swin Transformer Block ×2
Path Merging
Swin Transformer Block ×2
Path Merging
Swin Transformer Block ×6
Path Partition
Swin Transformer Block ×6
Stage 1
Stage 2
Stage 3
Stage 4
Backbone of Swin Transformer
Feature Map
36
1×1
Im_info
18
3×3
1×1
Proposal
Reshape
Softmax
Reshape
RPN
ROI
Softmax
Classification
Bbox_pred

Figure 4. C-SwinFaster Network structure diagram.

Faster R-CNN, on the other hand, optimizes calculations within each candidate region of a regional convolutional neural network. It passes features from the Region Proposal Network (RPN) to the last layer of the convolutional feature map and employs a dedicated ROI pooling layer. This approach allows simultaneous training of classification and boundary regression. While it addresses limitations in candidate box numbers found in RCNN and Fast RCNN, its commonly used backbone networks like ResNet and VGGNet fall short in identifying detailed features of crab juveniles' abdomens, impacting classification results. In contrast, the Transformer model, with its focus on pixel correlations, captures more subtle visual information, improving recognition rates for male and female crab characteristics. In addition, its self-attention mechanism has excellent global modeling capabilities, allowing the model to understand the image globally and understand the relationship between various parts of the image through the self-attention mechanism, thereby capturing the subtle differences in characteristics of male and female crabs.

In conclusion, the Swin Transformer not only demonstrates robust performance but also offers unique advantages in crab gender identification, making it an ideal choice as the backbone network for this study.

2.3.2. PAFPN

The Feature Pyramid Network (FPN) is the basic structure of the target detection framework and is widely used in multi-scale fusion [27]. The research subject of this article, Chinese mitten crab juveniles, presents challenges such as small size, similar shapes, and complex backgrounds. In the conventional FPN structure, different levels of feature maps convey varying degrees of information. Low-level features usually contain more edge information, while high-level features contain more global semantic information. This leads to an imbalance of information between feature maps, making it difficult to effectively fuse and transfer them. It reduces the number of channels of the feature map through a 1 × 1 convolution operation for feature fusion. However, this process may result in partial loss of semantic information due to the compression of feature channels. At the same time, FPN's feature map fusion mainly focuses on local areas and lacks global attention to other levels of feature maps. This may introduce noise during feature map fusion, thus affecting detection performance. Given this, this article proposes an improved feature pyramid network—Path Aggregation Feature Pyramid Network (PAFPN) combined with Swim Transformer. Different from traditional FPN, the core advantage of Path Aggregation Network (PAN) is that it can effectively aggregate features at different levels to provide a more comprehensive and continuous information flow. By combining this advantage of PAN with FPN, PAFPN can not only reduce feature loss but also enhance the model's detection ability for small, similar-shaped targets with complex backgrounds such as Chinese mitten crab larvae.

The PAFPN in this research combines a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN). As shown in Figure 5, the input image of a crab undergoes FPN's bottom-up feedforward processing to generate features {C2, C3, C4, C5}. This process transitions the information flow from low-level features, like the edge and color of the crab's abdomen, to high-level features, such as the overall part of the abdomen. Each stage involves down-sampling, resulting in subsequent feature maps that are higher in semantic level but smaller in size, which potentially leads to the loss of detailed information. Then, FPN performs top-down path enhancement, starting from the highest-level feature map, up-sampling it, and fusing it with the next-level feature map. This process repeats at each stage, ultimately yielding multi-scale feature maps rich in semantic information {P2, P3, P4, P5}. To better preserve the shallow feature information, this paper introduces bottom-up path aggregation, as illustrated by the dotted arrows. Shallow features first reach the P2 layer through lateral connections in the original FPN, then ascend to the top layer via bottom-up path augmentation. This approach enables feature maps at different levels to share information, undergo multiple iterations of bottom-up and top-down collection and fusion, and eventually form the fused {N2, N3, N4, N5} features for subsequent classification and regression tasks.

The detailed structure of bottom-up path augmentation is shown in Figure 6. After a convolution with a size of 3 × 3 and a stride of 2, the feature map size is reduced to half of the original size, and then compared with P_{i+1} The feature map is tensor connected, and the result is passed through a convolution layer with a convolution kernel size of 3 × 3 and a stride of 1 to obtain N_{i+1}.

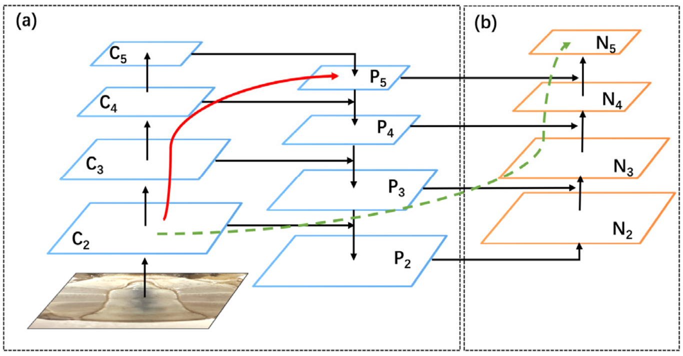

Figure 5. Path aggregation feature pyramid network structure diagram. (**a**) FPN backbone. (**b**) bottom-up path augmentation.

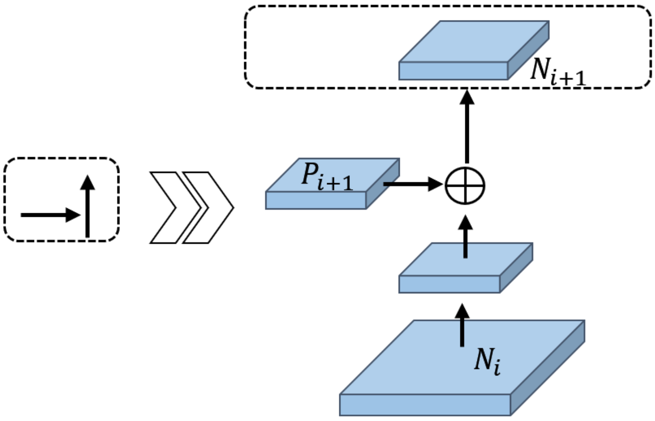

Figure 6. Illustration of building block of bottom-up path augmentation.

2.3.3. Softer-NMS

In the Faster R-CNN model, the Non-Maximum Suppression (NMS) algorithm is typically used to filter candidate boxes and anchor boxes [28]. NMS essentially identifies the most accurate bounding boxes for the target object. However, when training the model in the gender-characteristic area of crab juveniles, the characteristics of female crabs may include the characteristics of male crabs. As a result, as long as one round of gender misjudgment is female, the accuracy of male crab identification will be greatly reduced. Therefore, to avoid the accidental deletion and error detection problems caused by the traditional NMS algorithm, this article adopts the Soft-NMS algorithm. For boxes whose Intersection over Union (IOU) with the highest score box is greater than the threshold, they are not removed directly, but their confidence is reduced by increasing the weight associated with the IOU. This can retain more boxes and avoid accidental deletion to a certain extent. To avoid threshold setting, Gaussian weighting is used, as shown in Equation (1).

$$S_i = S_i e^{-\frac{iou(M,b_i)^2}{\sigma}}, \quad \forall b_i \notin D \tag{1}$$

In the formula: S_i is the score corresponding to the i-th prediction box; M is the candidate box with the highest score; b_i is the prediction box to be detected; IOU is the intersection ratio of M and b_i; σ is the penalty coefficient.

The specific steps of the Soft-NMS algorithm are: 1. Initialization: Remove all candidate boxes with scores lower than the preset threshold and mark the remaining candidate boxes as active. 2. Sorting: Sort all active candidate boxes by a score from high to low. 3. Candidate box selection: Select the current active candidate box with the highest score and add it to the selection list. 4. IOU calculation: Calculate the IOU value between the just selected candidate box and all other candidate boxes that are still active. 5. Update score: Update the score of each active candidate box and selected box based on its IOU value. 6. Check the end condition: If all candidate boxes have been checked, or there are no active candidate boxes left, terminate the algorithm; otherwise, return to the third step.

The effect of filtering candidate frames by NMS and Soft-NMS is shown in Figure 7. When using NMS to filter suggestion boxes, if the threshold is set too low and the filtering conditions are not strict, it may lead to false detections. The use of the Soft-NMS algorithm can solve the problem of insufficiently accurate filtering of suggestion boxes to a certain extent, thereby improving the recognition effect of the network model.

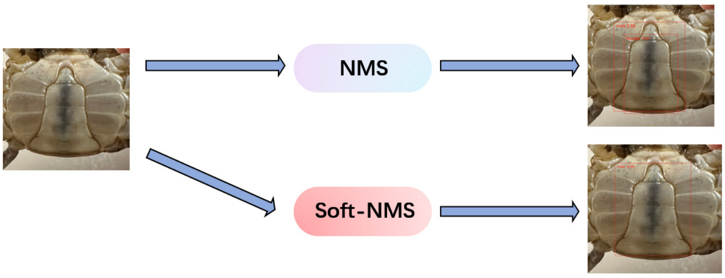

Figure 7. Comparison of the effects of improved NMS.

2.3.4. FocalLoss

Choosing an appropriate loss function is crucial. Currently, the binary cross-entropy loss function is widely used in segmentation, classification, and object detection tasks [29]. For the binary cross-entropy loss function, its definition is as follows:

$$L = -y\log y' - (1-y)\log(1-y') = \begin{cases} -\log y' & y = 1 \\ -\log(1-y') & y = 0 \end{cases} \tag{2}$$

In the formula: L is the cross entropy loss. When the sample is positive, $y = 1$, when it is negative, $y = 0$; and y' represents the probability value of a certain category output by the network, and the value range is between 0 and 1. The output of the cross-entropy loss function depends on the model's prediction of the probability value. For positive samples, a larger probability value indicates a prediction closer to the true value, decreasing the cross-entropy loss. On the contrary, for negative samples, the smaller the target probability value, the value of the loss function also decreases. This probability value is affected by each sample point. In the process of predicting the sex of crab juveniles, the existence of outliers or noise may confuse the model and reduce its prediction effect. At present, in the gender determination of Chinese mitten crab juveniles, many uncertainties and complex background factors may affect the final gender result, which increases the difficulty of using the cross-entropy loss function. Therefore, this study introduces the focal loss function to replace cross-entropy loss to further suppress the interference caused by complex environments. After adding parameters α and β in Equation (3), the calculation of focal loss is shown in Equation (3).

$$L_{\mathrm{F}} = \begin{cases} -\alpha(1-y')^{\beta}\log y' & y = 1 \\ -(1-\alpha)y'^{\beta}\log(1-y') & y = 0 \end{cases} \tag{3}$$

Returning to our task, which is gender classification, y =1 means "male" and y =0 means "female". The "α" factor is used to balance the weight of positive and negative samples ($0 < \alpha < 1$) and reduce the weight ratio of a large number of negative samples; while the "β" factor reduces the loss of easily classified samples. When $(1 - y')$ tends to 1 after being raised to the power of β, resulting in a smaller value of the loss function. However, the loss for positive samples with smaller probability values becomes relatively larger. The same goes for negative samples. By comparing the cross-entropy loss function and focal loss, it can be seen that focal loss can adjust the weight of easy-to-classify samples and difficult-to-classify samples in the loss function during model training. For our task, we can adjust the "α" value in focal loss to solve the problem of imbalanced samples of crab larvae of different genders. Increasing the value of "β" allows the model to pay more attention to samples that are difficult to distinguish. In the gender identification of male and female crab juveniles, these difficult-to-distinguish samples may be crucial for gender prediction. After multiple experiments and comparisons, it was found that when $\alpha = 0.25$ and $\beta = 2$, the loss is at its minimum.

2.3.5. ROI Integrating Attention Mechanism

To address the computational and memory challenges in the region proposal network (RPN) and region-based convolutional neural network (R-CNN) for processing extensive datasets, the concept of a Region of Interest (ROI) pooling network was introduced. This approach, designed to effectively extract features from specified regions of interest, significantly reduces computational complexity and enhances the efficiency of object detection tasks [30]. However, a notable limitation of ROI pooling networks is their inability to capture long-range dependencies and contextual information, which are vital for the precise and robust identification of Chinese mitten crab juveniles.

To solve these problems, this paper introduces a self-attention mechanism [31] into the ROI pooling network, whose structure is shown in Figure 8. First, the input value is transformed through three different convolutional layers to obtain the corresponding query, key, and value. These three transformations are used to construct query, key, and value information in the self-attention mechanism, respectively. Next, by performing a product operation on the query and key, and then applying the Softmax function to normalize the product result, the attention weights a1, a2, a3, and a4 are obtained. These weights are used for the weighted sum of values to form the output features. Finally, the attention output feature is residually connected to the input x and multiplied by the learnable parameter gamma to obtain the final output out.

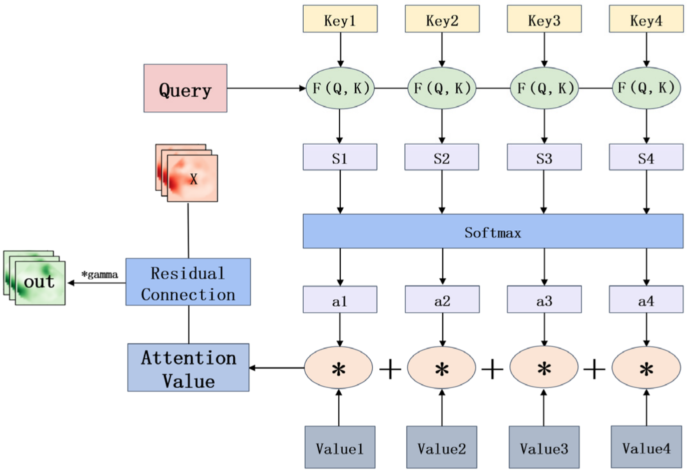

Figure 8. Illustration of the self-attention mechanism.

By integrating the self-attention mechanism, the model can assign varied importance levels to different regions based on their contextual relevance and interconnectedness. This capability allows for more effective capturing of the global context and spatial relationships among the crab image's detailed parts, significantly improving recognition and detection accuracy.

3. Experiment and Analysis

3.1. Experimental Configuration

In response to the critical need for robust and unbiased evaluation of model performance, this study incorporates the K-fold cross-validation technique, a method widely acknowledged for its efficacy in reducing biases associated with the random partitioning of data. For our analysis, we chose a five-fold cross-validation scheme, which repeats five different cross-validations and then averages the results of the 5 tests. This scheme was found to be optimal for our dataset size and complexity. This approach enables a more thorough investigation into the model's ability to generalize, thereby ensuring the robustness and reliability of our findings. By implementing five-fold cross-validation, we mitigate the risk of overfitting and provide a more accurate representation of the model's effectiveness in real-world scenarios.

For implementing the methodology of this paper, our experimental procedure was powered by the PyTorch framework in conjunction with the mmdetection open-source tool, provided by the University of Hong Kong. This was executed on a system operating with Ubuntu 20.04.4. The hardware specification for our project included an Intel Core i9-11900KF CPU, 64 GB RAM, and an NVIDIA 3090 GPU (24 GB). The software environment was primarily composed of PyTorch 1.6.0, Python 3.8, CUDA 11.7, and cuDNN 8.5.0. Together, these elements provided a robust platform for our computational and analytical needs during the experiment. This configuration ensured effective resource management and efficient code execution.

3.2. Evaluation Metrics

To quantitatively analyze the effect of this algorithm in target detection, we use mean average precision (mAP) to evaluate the accuracy of the model.

Among the evaluation indicators, we divide the judgment results into the 4 categories listed in Table 2 based on the detection results and the actual situation of the image.

Table 2. Discrimination result classification.

Classification Result Category	**Actual Sample Category**	
	Positive Sample	**Negative Sample**
Positive Sample	True Positive (TP)	False Positive (FP)
Negative Sample	False Negative (FN)	True Negative (TN)

Precision, that is, the accuracy rate, refers to the proportion of correct samples among all recognized images, as shown in Equation (4):

$$p = \frac{TP}{TP + FP} \tag{4}$$

Recall, or recall rate, refers to the rate of correct identification among all labelled positive samples, as shown in Equation (5).

$$r = \frac{TP}{TP{+}FN} \tag{5}$$

The average precision of a single category is drawn with r as the x-axis and p as the y-axis, and the P-R curve is drawn. The AP value is calculated from the area under the curve:

$$AP = \int_0^1 p(r)dr \tag{6}$$

mAP refers to the average AP value under all categories, which is an important indicator for judging the performance of the target detection algorithm. For a test set of N categories, the mAP calculation formula is as follows:

$$mAP = \frac{\sum_{k=1}^{N} AP}{N} \tag{7}$$

3.3. Network Training Results

During the training phase, our model, referred to as C-SwinFaster, exhibits a significant learning trajectory, as evidenced by the loss curve depicted in Figure 9. The curve initiates from a high value, indicative of the model's initial unfamiliarity with the task. As training progresses, the loss decreases rapidly, reflecting the model's growing proficiency in learning from the data. This rapid decline is most pronounced in the initial 10,000 iterations, a phase characterized by the model's acquisition of fundamental features and patterns from the dataset. After this point, the curve begins to stabilize, signifying that the model is approaching convergence.

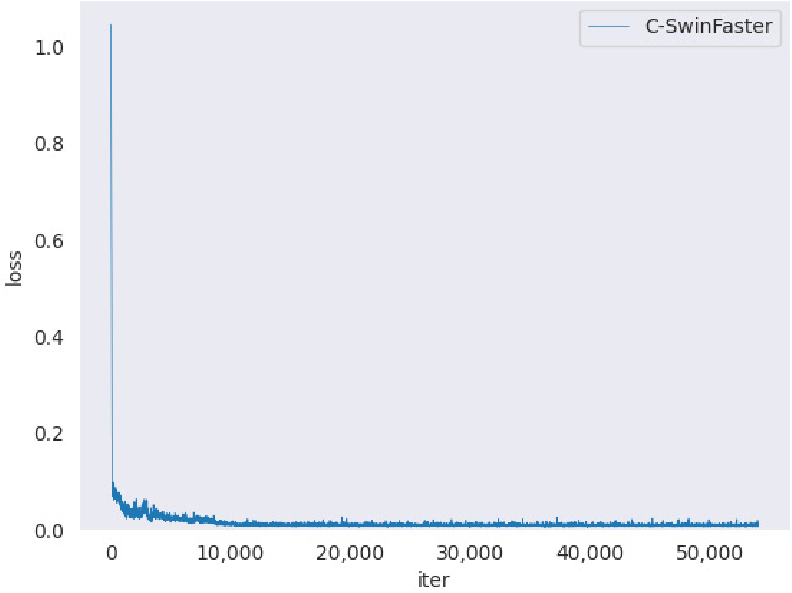

Figure 9. Illustration of training loss curves of C-SwinFaster.

A crucial aspect of the training phase is the absence of significant spikes in the loss curve, which typically signal overfitting. Overfitting occurs when a model becomes too tailored to the training data, losing its generalizability to new, unseen data. The smooth nature of the C-SwinFaster model's loss curve suggests that it has maintained a balance between learning the intricacies of the training dataset and retaining the flexibility to perform well on new data.

Remarkably, the loss function maintains a consistent downward trajectory throughout the training process, with minor fluctuations indicative of the model's adaptation to more nuanced patterns and anomalies within the training dataset. The loss approaching zero after 2000 iterations is a strong indicator of the model's high accuracy in prediction. This suggests that the C-SwinFaster model has effectively learned the distinguishing features necessary for accurately identifying the sex of Chinese mitten crab juveniles.

The PR curve is a graphical representation that elucidates the trade-off between precision, the proportion of true positive results among all positive identifications, and recall, the proportion of true positive results among all relevant samples. Figure 10 presents the Precision-Recall (PR) curve for the C-SwinFaster model. It is obvious from the curve that the C-SwinFaster model maintains high accuracy (above 0.8) at most recall thresholds. Precision values drop significantly as recall approaches 1.0, which is typical as trying to catch all positive samples (recall = 1.0) usually results in more false positives (lower precision). The sharp drop on the far right side of the chart indicates that the model is forced to compromise precision to a greater extent to achieve the highest recall.

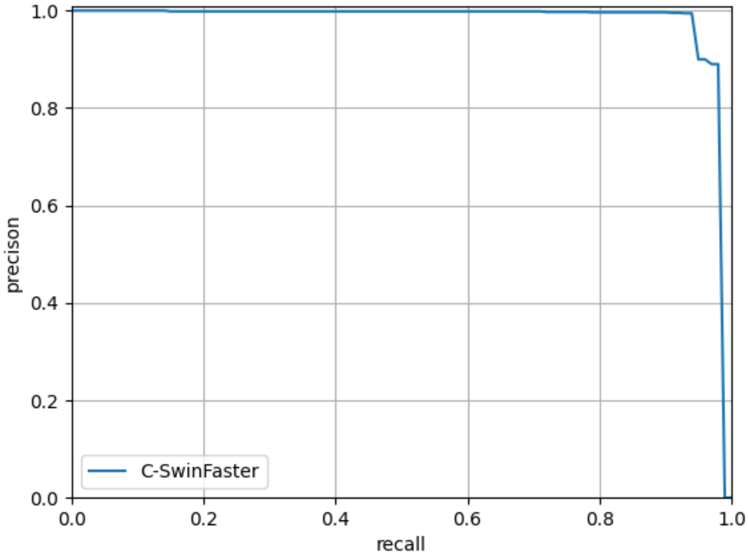

Figure 10. Illustration of PR curve.

This behavior demonstrates that the C-SwinFaster model is very robust in the task of classifying the sex of Chinese mitten crab juveniles, with a strong performance in both precision and recall over most of its operating range. This shows that the model is effective in correctly identifying positive samples while minimizing the number of false positives until high recall levels are achieved.

4. Discussion

4.1. Comparison of Different Algorithms

To enhance the effectiveness and scientific rigor of the algorithms in this paper, but based on the fact that there is no relevant algorithm as well as improved algorithms applied to the sex recognition of Chinese mitten crab juveniles at this stage for the time being, we had to choose more popular target detection algorithms on underwater organisms in recent years for comparative analysis, including Faster R-CNN, Cascade-RCNN, Mask RCNN, SSD and YOLOv8. Uniform training parameters were maintained for all experiments, and the results are detailed in Table 3.

Table 3. Comparison of the algorithm in this paper with other algorithms.

Algorithm	Backbone	mAP50-95 (%)	mAP50 (%)	mAP75 (%)
Faster R-CNN	ResNet-50	88.12	89.11	86.53
Cascade-RCNN	ResNet-101	92.73	93.76	90.87
Mask RCNN	ResNet-101	89.22	91.64	90.59
SSD	SSDVGG	86.69	88.77	84.98
YOLOv8	CSPDarkNet	95.54	96.01	92.36
C-SwinFaster	Swin Transformer	98.45	98.79	98.19

The comparative data illustrates that, on the crab juveniles dataset, the mAP50-95 of our C-SwinFaster algorithm exhibited notable improvements: 10.33% over Faster R-CNN, 5.72% over Cascade-RCNN, 9.23% over Mask RCNN, 11.76% over SSD, and 2.91% over YOLOv8, respectively. The model's performance was particularly distinguished at an Intersection Over Union (IOU) of 75, where its mAP significantly surpassed that of the other models. This underscores the efficacy of the C-SwinFaster approach in the intricate task of sex identification in Chinese mitten crab juveniles.

Figure 11 graphically represents the mAP trends of each network model over 50 training cycles. The C-SwinFaster network rapidly achieved superior mAP values. It exhibited a swift ascent in the initial epochs, subsequently maintaining the highest performance plateau, outperforming the other models under evaluation.

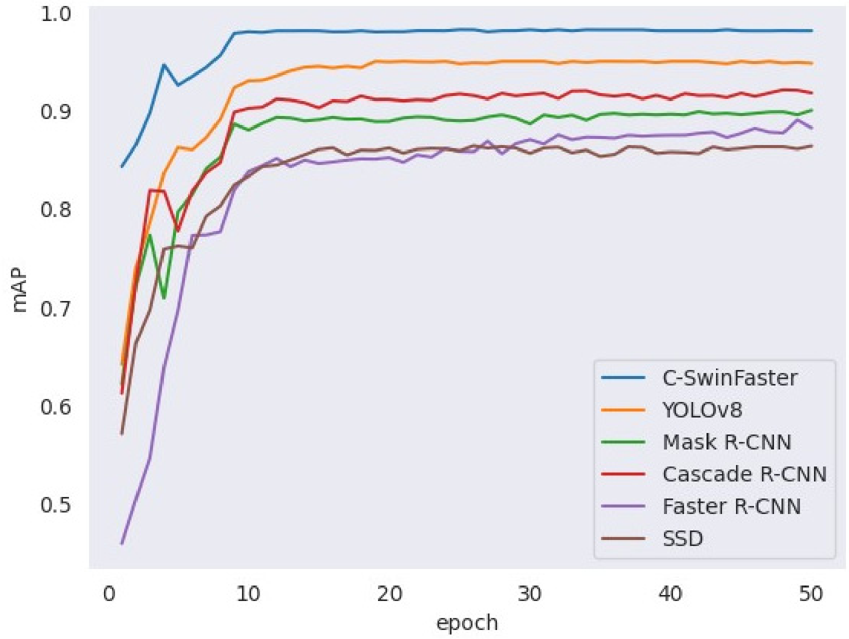

Figure 11. Illustration of mAP curves of the algorithm in this paper with other algorithms.

4.2. Ablation Experiment

Ablation experiments are crucial in scientific research, primarily to analyze the impact of different network branches on the overall model performance. To evaluate the optimization effect and efficiency of each module in our backbone feature extraction network, we conducted a series of ablation studies on the following components: Swin Transformer, PAFPN, Soft-NMS, Focal Loss, and the RoI pooling network integrated with an attention mechanism. The detailed experimental settings and results are presented in Table 4, with a '√' indicating the incorporation of these enhancements.

Table 4. Results of the ablation experiments.

Method	**Swin-T**	**PAFPN**	**Soft-NMS**	**Focal Loss**	**Attention**	**Time (s)**	**mAP (%)**
Faster R-CNN						31	88.12
Improve1	√					41	90.88
Improve2	√	√				35	92.67
Improve3	√	√	√			32	93.98
Improve4	√	√	√	√		28	94.69
C-SwinFaster	√	√	√	√	√	33	98.45

The table delineates the analysis across six experimental groups, focusing on two primary indicators: mAP (mean Average Precision) and training time. Under consistent training parameters, the mAP of the baseline Faster R-CNN model was recorded at 88.12%, with a processing time of 31 s for 160 test images. The integration of the Swin Transformer as the backbone feature extraction network resulted in a 2.76% increase in mAP; however, this enhancement led to a 10-s increase in processing time due to heightened network complexity. The addition of PAFPN simplified the overall network complexity, further raising the mAP by 1.79% and reducing the processing time by 6 s. Implementing the Soft-NMS algorithm and Focal Loss function augmented the mAP by 5.86% and 6.57%, respectively, while significantly curtailing the test time. Ultimately, the C-SwinFaster algorithm, despite requiring marginally more testing time, achieved a substantial 10.33%

increase in mAP compared to the traditional Faster R-CNN algorithm. These results affirm the effectiveness of the proposed strategies in this study, particularly in the sex identification of male and female crab juveniles.

For a more intuitive assessment of the improvements, the pre-and post-improvement models were tested on a subset of the validation set. The outcomes are depicted in Figure 12. The figure illustrates the detection results of the standard Faster R-CNN algorithm and the enhanced C-SwinFaster algorithm on identical datasets, specifically (G_1)–(G_4) and (H_1)–(H_4). It is evident that the improvements have led to increased overall network accuracy, with notable enhancements in reducing missed and false detections.

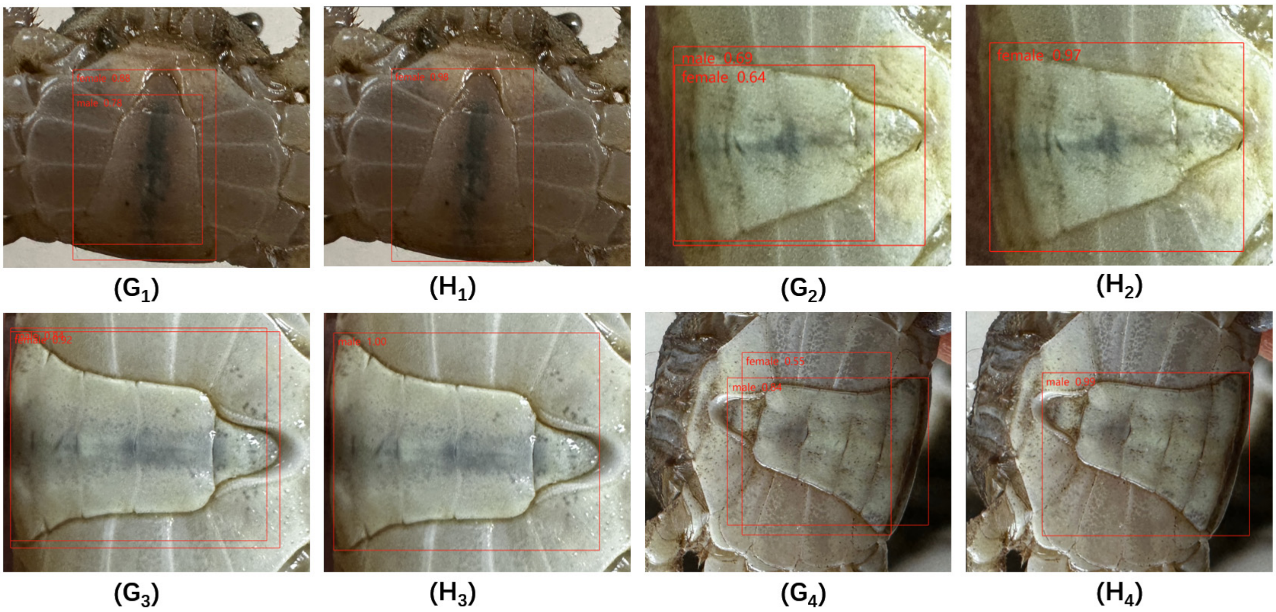

Figure 12. Illustration of the results of the improved algorithm in this article.

5. Conclusions

In this study, we have merged deep learning technology with the unique needs of Chinese mitten crab juveniles to intelligently differentiate between male and female specimens. This advancement fosters the transition of modern fishery sorting equipment from semi-mechanized systems to intelligent ones. We introduce an improved Faster R-CNN algorithm. By adjusting image characteristics like brightness, adding noise, rotating, cropping, flipping, and employing other data augmentation techniques, we effectively simulated a variety of real-world conditions relevant to practical applications. The integration of the Swin Transformer and the path aggregation feature pyramid network has greatly enhanced the extraction of complex features pertinent to this task. Implementing soft non-maximum suppression and focal loss markedly improved overall detection accuracy. Additionally, integrating a region-of-interest pooling network with a self-attention mechanism enhanced the model's ability to understand both intra- and inter-regional correlations, thereby improving feature representation. This multifaceted enhancement not only improved accuracy but also expedited processing speed. The refined model attained a mAP of 98.45%. The application of this advanced algorithm holds the potential to elevate the level of automation in aquaculture, offering substantial labor cost reductions and significantly boosting the efficiency of large-scale Chinese mitten crab farming. This study serves as a valuable reference for propelling forward smart aquaculture practices.

Looking ahead, the potential applications of this research extend beyond the realm of aquaculture. The methodologies and technologies developed here can be adapted to other species and industries, paving the way for broader impacts in automated classification and environmental sustainability. However, challenges such as data availability, model generalizability, and real-world application constraints remain. Future research should focus on these areas, aiming to enhance the robustness and applicability of our model across diverse scenarios. Moreover, this research contributes to the technological advancements in machine learning and computer vision. By successfully applying the Swin Transformer to a novel domain, we demonstrate its versatility and potential for further innovative applications. This study also underscores the importance of interdisciplinary approaches, combining expertise in biology, computing, and engineering, to address complex real-world problems.

In summary, our study not only achieves high accuracy in the gender identification of Chinese mitten crab juveniles but also opens avenues for future research and application in various fields. It stands as a testament to the power of combining cutting-edge technology with practical applications, driving forward the progress in smart aquaculture and beyond.

Author Contributions: Conceptualization, M.C.; methodology, H.G.; software, H.G.; validation, D.G.; resources, M.C.; data curation, D.G.; writing—original draft, H.G.; writing—review and editing, M.C.; visualization, D.G.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research and Development Planning in Key Areas of Guangdong Province, NO. 2021B0202070001.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to Significant investment of time and money in data acquisition.

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

References

1. Xu, G.; Han, W.; Sun, Y.; Lu, Z.; Long, Q.; Cheng, Y. Impacts of Three Feeding Modes on the Diet Composition and Contribution Ratios for Juvenile Chinese Mitten Crab *Eriocheir sinensis*. *Aquac. Rep.* **2023**, *29*, 101516. [CrossRef]
2. Zhu, S.; Long, X.; Turchini, G.M.; Deng, D.; Cheng, Y.; Wu, X. Towards Defining Optimal Dietary Protein Levels for Male and Female Sub-Adult Chinese Mitten Crab, *Eriocheir sinensis* Reared in Earthen Ponds: Performances, Nutrient Composition and Metabolism, Antioxidant Capacity and Immunity. *Aquaculture* **2021**, *536*, 736442. [CrossRef]
3. Liu, X.; Wu, H.; Wang, Y.; Liu, Y.; Zhu, H.; Li, Z.; Shan, P.; Yuan, Z. Comparative Assessment of Chinese Mitten Crab Aquaculture in China: Spatiotemporal Changes and Trade-Offs. *Environ. Pollut.* **2023**, *337*, 122544. [CrossRef] [PubMed]
4. Zhang, G.; Jiang, X.; Zhou, W.; Chen, W.; Levy, T.; Wu, X. Stocking Density Affects Culture Performance and Economic Profit of Adult All-Female Chinese Mitten Crabs (*Eriocheir sinensis*) Reared in Earthen Ponds. *Aquaculture* **2024**, *581*, 740352. [CrossRef]
5. Yang, Z.; Wei, B.; Liu, Q.; Cheng, Y.; Zhou, J. Individual Growth Pattern of Juvenile Stages of the Chinese Mitten Crab (*Eriocheir sinensis*) Reared under Laboratory Conditions. *Aquac. Int.* **2018**, *26*, 645–657. [CrossRef]
6. Zhang, J.; Wang, S.; Zhang, S.; Li, J.; Sun, Y. Research on Target Detection and Recognition Algorithm of *Eriocheir sinensis* Carapace. *Multimed. Tools Appl.* **2023**, *82*, 42527–42543. [CrossRef]
7. Liu, C.; Wang, Z.; Li, Y.; Zhang, Z.; Li, J.; Xu, C.; Du, R.; Li, D.; Duan, Q. Research Progress of Computer Vision Technology in Abnormal Fish Detection. *Aquac. Eng.* **2023**, *103*, 102350. [CrossRef]
8. Liu, C.; Liu, Y.; Zhang, Q.; Li, X.; Wu, T.; Li, Q. A Two-Stage Classification Algorithm for Radar Targets Based on Compressive Detection. *EURASIP J. Adv. Signal Process* **2021**, *2021*, 23. [CrossRef]
9. Xu, G.; Xu, X.; Gao, H.; Xiao, F. FP-RCNN: A Real-Time 3D Target Detection Model Based on Multiple Foreground Point Sampling for Autonomous Driving. *Mob. Netw. Appl.* **2023**, *28*, 369–381. [CrossRef]
10. Ahmad, M.; Ahmed, I.; Jeon, G. An IoT-Enabled Real-Time Overhead View Person Detection System Based on Cascade-RCNN and Transfer Learning. *J. Real. Time Image Process* **2021**, *18*, 1129–1139. [CrossRef]
11. Arora, N.; Kumar, Y.; Karkra, R.; Kumar, M. Automatic Vehicle Detection System in Different Environment Conditions Using Fast R-CNN. *Multimed. Tools Appl.* **2022**, *81*, 18715–18735. [CrossRef]
12. Zhang, Z.; Shi, R.; Xing, Z.; Guo, Q.; Zeng, C. Improved Faster Region-Based Convolutional Neural Networks (R-CNN) Model Based on Split Attention for the Detection of Safflower Filaments in Natural Environments. *Agronomy* **2023**, *13*, 2596. [CrossRef]
13. Zhang, L.; Wang, H.; Wang, X.; Liu, Q.; Wang, H.; Wang, H. Vehicle Object Detection Method Based on Candidate Region Aggregation. *Pattern Anal. Appl.* **2021**, *24*, 1635–1647. [CrossRef]
14. Qiu, Z.; Wang, S.; Zeng, Z.; Yu, D. Automatic Visual Defects Inspection of Wind Turbine Blades via YOLO-Based Small Object Detection Approach. *J. Electron. Imaging* **2019**, *28*, 43023. [CrossRef]
15. Ding, L.; Xu, X.; Cao, Y.; Zhai, G.; Yang, F.; Qian, L. Detection and Tracking of Infrared Small Target by Jointly Using SSD and Pipeline Filter. *Digit. Signal Process* **2021**, *110*, 102949. [CrossRef]

16. Mu, X.; He, L.; Heinemann, P.; Schupp, J.; Karkee, M. Mask R-CNN Based Apple Flower Detection and King Flower Identification for Precision Pollination. *Smart Agric. Technol.* **2023**, *4*, 100151. [CrossRef]
17. Cao, S.; Zhao, D.; Sun, Y.; Ruan, C. Learning-Based Low-Illumination Image Enhancer for Underwater Live Crab Detection. *ICES J. Mar. Sci.* **2021**, *78*, 979–993. [CrossRef]
18. Li, Y.; Zhang, X.; Shen, Z. YOLO-Submarine Cable: An Improved YOLO-V3 Network for Object Detection on Submarine Cable Images. *J. Mar. Sci. Eng.* **2022**, *10*, 1143. [CrossRef]
19. Zhang, J.; Shi, Y.; Yang, J.; Guo, Q. KD-SCFNet: Towards More Accurate and Lightweight Salient Object Detection via Knowledge Distillation. *Neurocomputing* **2024**, *572*, 127206. [CrossRef]
20. Ji, W.; Peng, J.; Xu, B.; Zhang, T. Real-Time Detection of Underwater River Crab Based on Multi-Scale Pyramid Fusion Image Enhancement and MobileCenterNet Model. *Comput. Electron. Agric.* **2023**, *204*, 107522. [CrossRef]
21. Liu, S.; Zhang, X.; Wang, X.; Hou, X.; Chen, X.; Xu, J. Tomato Flower Pollination Features Recognition Based on Binocular Gray Value-Deformation Coupled Template Matching. *Comput. Electron. Agric.* **2023**, *214*, 108345. [CrossRef]
22. Cui, Y.; Pan, T.; Chen, S.; Zou, X. A Gender Classification Method for Chinese Mitten Crab Using Deep Convolutional Neural Network. *Multimed. Tools Appl.* **2020**, *79*, 7669–7684. [CrossRef]
23. Xue, J.; Jiang, T.; Chen, X.; Liu, H.; Yang, J. Multi-Mineral Fingerprinting Analysis of the Chinese Mitten Crab (*Eriocheir sinensis*) in Yangcheng Lake during the Year-Round Culture Period. *Food Chem.* **2022**, *390*, 133167. [CrossRef]
24. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
25. Su, P.; Han, H.; Liu, M.; Yang, T.; Liu, S. MOD-YOLO: Rethinking the YOLO Architecture at the Level of Feature Information and Applying It to Crack Detection. *Expert Syst. Appl.* **2024**, *237*, 121346. [CrossRef]
26. Wang, J.; Zeng, Z.; Sharma, P.K.; Alfarraj, O.; Tolba, A.; Zhang, J.; Wang, L. Dual-Path Network Combining CNN and Transformer for Pavement Crack Segmentation. *Autom. Constr.* **2024**, *158*, 105217. [CrossRef]
27. Xu, Z.; Li, T.; Liu, Y.; Zhan, Y.; Chen, J.; Lukasiewicz, T. PAC-Net: Multi-Pathway FPN with Position Attention Guided Connections and Vertex Distance IoU for 3D Medical Image Detection. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1049555. [CrossRef]
28. Ding, J.; Zhang, J.; Zhan, Z.; Tang, X.; Wang, X. A Precision Efficient Method for Collapsed Building Detection in Post-Earthquake UAV Images Based on the Improved NMS Algorithm and Faster R-CNN. *Remote Sens.* **2022**, *14*, 663. [CrossRef]
29. Hu, K.; Zhang, Z.; Niu, X.; Zhang, Y.; Cao, C.; Xiao, F.; Gao, X. Retinal Vessel Segmentation of Color Fundus Images Using Multiscale Convolutional Neural Network with an Improved Cross-Entropy Loss Function. *Neurocomputing* **2018**, *309*, 179–191. [CrossRef]
30. Priyanka; Baranwal, N.; Singh, K.N.; Singh, A.K. YOLO-Based ROI Selection for Joint Encryption and Compression of Medical Images with Reconstruction through Super-Resolution Network. *Future Gener. Comput. Syst.* **2024**, *150*, 1–9. [CrossRef]
31. Hou, P.; Zhang, J.; Jiang, Z.; Tang, Y.; Lin, Y. A Bearing Fault Diagnosis Method Based on Dilated Convolution and Multi-Head Self-Attention Mechanism. *Appl. Sci.* **2023**, *13*, 2770. [CrossRef]

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.