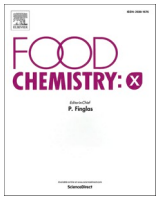
Food Chemistry: X 19 (2023) 100733

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Enhanced batch sorting and rapid sensory analysis of Mackerel products using YOLOv5s algorithm and CBAM: Validation through TPA, colorimeter, and PLSR analysis

Yi-Zhen Huang [a,b], Lin Han [a,b], Xiaoqing Yang [a,b], Yu Liu [a,b], Bei-Wei Zhu [a,b,*], Xiu-Ping Dong [a,b,*]

[a] Academy of Food Interdisciplinary Science, School of Food Science and Technology, Dalian Polytechnic University, Dalian 116034, Liaoning, China
[b] National Engineering Research Center of Seafood, Collaborative Innovation Center of Seafood Deep Processing, Liaoning Province Collaborative Innovation Center for Marine Food Deep Processing, Dalian 116034, Liaoning, China

ARTICLE INFO

Keywords:
Mackerel
Rapid method
Deep learning
Object detection
Sensory

ABSTRACT

This study employed the YOLOv5s algorithm to establish a rapid quality identification model for Pacific chub mackerel (*S. japonicus*) and Spanish mackerel (*S. niphonius*). Data augmentation was conducted using the copy-paste augmentation within the YOLOv5s network. Furthermore, a small object detection layer was integrated into the network structure's neck, while the convolutional block attention module (CBAM) was incorporated into the convolutional module to optimize the model. The model's accuracy was assessed through sensory evaluation, texture profile analysis, and colorimeter analysis. The findings indicated that the enhanced model achieved a mAP@0.5 score of 0.966, surpassing the original version's score of 0.953. Moreover, the improved model's params was only 7.848 M, and an average detection time of 115 ms/image (image resolution 2400 × 3200). Furthermore, sensory and physicochemical indicators are reliably distinguished between qualified and unqualified samples. The PLSR model exhibited R^2X, R^2Y, and Q^2 values of 0.977, 0.956, and 0.663, respectively.

1. Introduction

Mackerel, a pelagic migratory fish in the same family as tuna, is an economically important species (EUMOFA, 2019). Due to their large catches and strong seasonality, mackerel are mostly caught in purse seines and are characterized by fat content and perishable characteristics. Thus, the detection of the storage quality of mackerel in large quantities and multiple batches is a challenging yet important task. Traditional methods for evaluating mackerel include sensory evaluation, physical and chemical analysis. Traditional sensory evaluation involves using human subjects as a tool to determine the sensory characteristics of the product. The advantage of this approach is that the results are closer to the actual consumer perception than the physical and chemical test indicators, and, therefore, it is considered one of the most important methods for fisheries and fish-inspection departments to quickly evaluate the quality of fish and fish products and perform sorting (Hyldig et al., 2012). However, sensory evaluation is influenced by factors such as sample size, variation in species, and evaluator experience; therefore, it is often combined with sampling based on instrumental testing to jointly determine product quality and product classification. These instrumental tests include colorimeter, TPA, mass spectrometer, and electronic nose, among others, and some fast, less destructive, and objective methods such as LF-NMR. However, for large-scale hauls of several tons, tens of tons, or more, the usual means of testing, even if based on sampling, require a great deal of time and resources.

In recent years, due to the powerful learning and generalization ability of deep learning, real-time evaluation of large-volume and multi-batch samples in food production has become possible. This can reduce the leakage of traditional sampling and analysis and make inspections more comprehensive. Deep learning (DL) is an aspect of machine-learning research which has been introduced to bring machine learning closer to achieving its original goal of artificial intelligence. Deep learning allows computational models consisting of multiple processing layers to learn data representations with multiple levels of abstraction. These methods have dramatically increased the level of

* Corresponding authors at: Academy of Food Interdisciplinary Science, School of Food Science and Technology, Dalian Polytechnic University, Dalian 116034, Liaoning, China.
E-mail addresses: zhubeiwei@163.com (B.-W. Zhu), dxiuping@163.com (X.-P. Dong).

https://doi.org/10.1016/j.fochx.2023.100733
Received 4 March 2023; Received in revised form 22 May 2023; Accepted 28 May 2023
Available online 1 June 2023

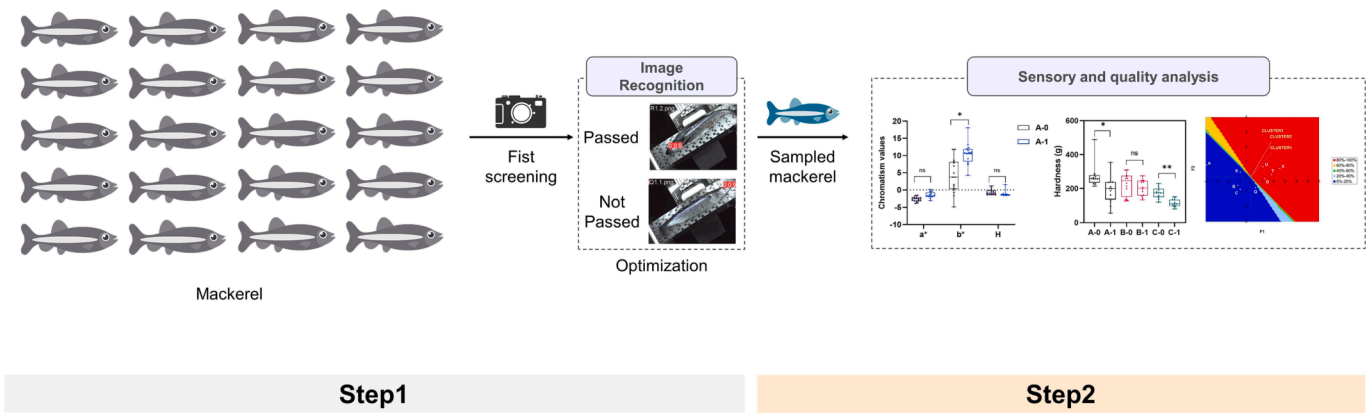


Fig. 1. Experiment design.

sophistication in speech recognition, visual-object recognition, object detection, and many other areas, such as drug discovery and genomics. Deep learning identifies complex structures in large datasets using backpropagation algorithms to indicate how a machine should change the internal parameters used to compute each layer of representation from the previous layer (LeCun et al., 2015). With the intensive research and development of deep-learning target-detection algorithms, their applications are increasing and they have begun to play a prominent role in improving the efficiency of the food industry. Zhou et al. (2019) surveyed dozens of articles in which deep learning is used to solve problems and adaptations in the food domain, including the quality inspection of fruits, vegetables, meat, and aquatic products; food supply chains; food identification; calorie estimation; and food contamination. Deep learning is now yielding promising results for classification and regression problems, attracting more research attention and making the real-time characterization of large batches of fish possible.

Current deep-learning-based target-detection methods can be divided into two categories. One is based on candidate region detection, such as R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), and Faster R-CNN (Ren et al., 2015). Another is based on regression detection, such as the YOLO series algorithm (Redmon et al., 2016). Essentially, the YOLO series algorithm first extracts the feature map of the image through a convolutional neural network, then inputs the feature map to the prediction network layer for prediction, and, finally, outputs the prediction result. The YOLO series algorithms are widely used for object detection because of their fast detection speed and high detection efficiency.

Considering that the real-time detection of large batches of mackerel has high requirements in terms of speed, the YOLO series algorithm is better able to meet its needs. It is worth noting that some tasks which are very simple for the human eye, such as identifying small tears and small color spots on the fish body, are very difficult for a computer that can only process the values 1 and 0. The YOLOv5 model uses the anchor sampling strategy to increase the number of positive samples, accelerate the network convergence, and reduce the difficulty of the regression of border parameters. Moreover, the YOLOv5 model performs data enhancement based on scaling, color space adjustment, and mosaic enhancement. Thus, the YOLOv5 model introduced in 2020 can be better adapted to practical engineering needs (Zhu et al., 2021). Furthermore, some data enhancement and attention mechanisms have been used in combination for deep-learning optimization. Ghiasi et al. (2012) conducted a systematic study for copy paste augmented and found that copy paste can provide reliable gains over a strong baseline. Sanghyun et al. (2018) innovatively proposed an attention mechanism that fuses channel attention with spatial attention to increase the representational power of the network by focusing on important features and suppressing unnecessary features. This study confirmed that the performance of various networks was greatly improved according to several benchmarks (ImageNet-1 K, MS COCO, and VOC, 2007) by inserting lightweight modules (CBAM).

In this study, a high-performance computer-aided intelligent quality-recognition technique based on the YOLO series algorithm is proposed. Using sample images directly as input data, fast and easy-to-understand results are provided to users. The primary object of this study was to meet the need for rapid sensory quality evaluation of Pacific chub mackerel and Spanish mackerel while supporting the quality control of high-volume samples and multi-batch sample storage. In order to validate the quality properties of the model judgments, we optimized the existing model and combined it with sensory evaluation, texture analysis, and color difference analysis. The incorporation of physical and sensory tests in our study serves several important purposes. Firstly, a comprehensive assessment of the mackerel samples was provided by considering not only the algorithm's output but also the sensory attributes, texture characteristics, and visual appearance. This holistic evaluation ensures that the model's performance aligns with human perception and preferences, which is essential for real-world application. Secondly, by integrating sensory evaluation, texture analysis, and color difference analysis, the research aimed to establish a multi-dimensional understanding of the mackerel samples' quality attributes. The sensory evaluation allows trained assessors to provide subjective judgments regarding taste, aroma, and overall preference, which are crucial factors influencing consumer acceptance. The texture analysis provides insights into the physical properties of the mackerel, such as springiness and hardness, which greatly impact the eating experience. Additionally, the color difference analysis objectively measures any changes in visual appearance, including potential discolorations or abnormalities. By combining these methods, this study aimed to provide an efficient and accurate approach for batch sorting and rapid sensory analysis of mackerel products, ultimately ensuring the delivery of high-quality and safe products to consumers. The inclusion of physical and sensory tests in the evaluation process adds valuable context, validity, and reliability to our proposed computer-aided technique.

2. Materials and methods

2.1. Experiment design

The experiment consisted of two parts. Step 1 was the primary screening of 1100 samples based on the small object-detection method of YOLOv5 to identify Pacific chub mackerel and Spanish mackerel pictures. Step 2 included the sensory evaluation and instrumental analysis, which were performed after the primary screening (Fig. 1).

Step 1: Picture acquisition for 1100 fish samples. Image data was input, using YOLOv5′s detection algorithm, into the mackerel image training model. Using the trained detection model, the relevant fish pictures were tested and the border output with the highest confidence

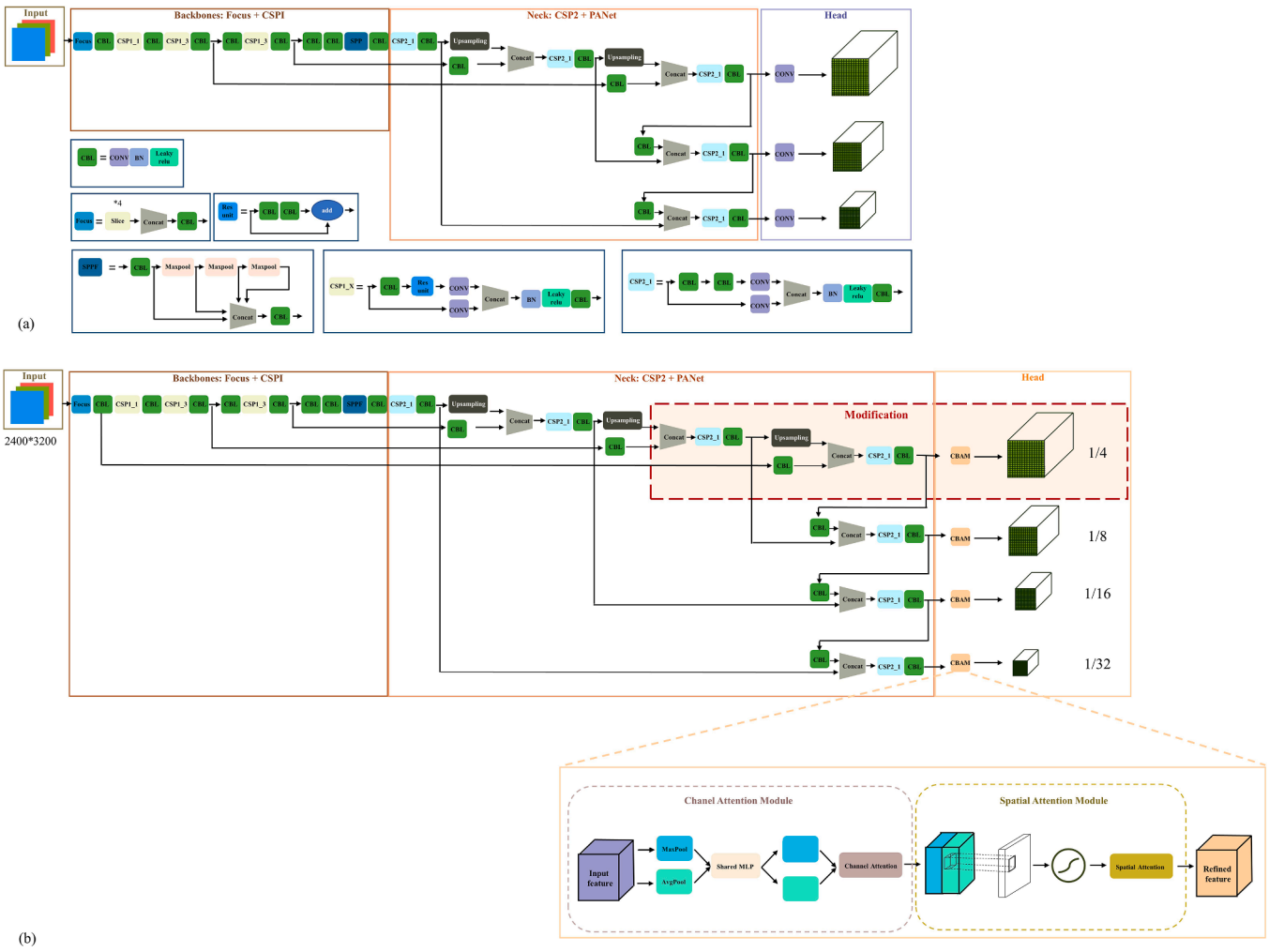


Fig. 2. Model structure. *Note.* (a) Network structure of YOLOv5s; (b) Network structure of improved YOLOv5s.

was selected to accomplish the goal of identifying mackerel with YOLOv5. The model output results in 1 and 0, 0 indicates that the sample is not the current year batch in the initial sieving of image recognition, and the sample is not passed. 1 indicates that the sample is the current year batch in the initial sieving of image recognition, and the sample is passed.

Step 2: For the output of the model, 36 samples from 1 and 36 samples from 0 were selected, respectively. These were divided into three groups according to the origin variety of the samples, with 12 samples in each group for color difference detection and texture detection, respectively. At the same time, nine additional samples were taken from each group and analyzed for sensory evaluation together with the 12 samples.

2.2. Raw material

Pacific chub mackerel, so-called Japanese mackerel (*S. japonicus*) (from Japan; labeled JP-M), Pacific chub mackerel (from Dalian, Yellow Sea; labeled SJ-M), and Spanish mackerel (*S. niphonius*) (from Dalian, Yellow Sea; labeled DS-M) were purchased from Yihexing Shandong Co. Ltd. The average weight and length were 430 g and 35 cm (Pacific chub mackerel), and 599 g and 45 cm (Spanish mackerel), respectively. Three different batches of fish samples (2019, 2020, 2021) were also purchased and all fish samples were stored in the Yihexing Shandong Co. Ltd. cold storage before purchase. The fish were then transferred in an insulated ice box via cold-chain transport (−4 °C) to the National Engineering Research Center of Seafood of Dalian Polytechnic University for storage at −20 °C.

2.3. Colorimeter analysis

Color changes were measured using a Konica-Minol-TA CR-400 colorimeter according to Thiansilakul et al. (2007) and visualized using the CIELab system. The location of the ventral part of the fish, 2 cm behind the gills and 1 cm below the fins, was selected for color variation detection. Six features were obtained to describe the color of the fish abdomen. The six features include L* (luminance), a* (redness/greenness), b* (yellowness/blueness), W (whiteness), ΔE (resolving degree), and H (hue).

2.5. Texture analysis

TPA analysis was performed with a texture analyzer (TA.XT.plus, Stable Micro Systems Ltd., UK) at 50% strain and 5 g trigger force with a

spherical plunger (5 mm diameter, P/50). The test parameters were as follows: pre-test, test, and return velocities of 1 mm/s, 1 mm/s, and 1 mm/s, respectively (Liu et al., 2021). Each test was performed 16 times.

2.6. Sensory evaluation

Training was conducted on the evaluation process according to ISO standards (2012); panelists were trained to identify Pacific chub mackerel and Spanish mackerel from different years through triangle tests; and, finally, the products were further evaluated by 21 panelists. The panelists valued the samples for pupil clarity, fish-belly color, freshness, and overall liking (on a 7-point scale) (Table S1). The assessors were given a short introduction to the whole experiment (including sample information). Despite the inability of our institution to provide an ethics committee, it is important to note that the fish samples utilized in this sensory experiment were natural food products with no additional ingredients. These samples were sold to consumers in many cities in China, and were confirmed to be safe for consumption.

All sensory evaluations were conducted in a standard sensory booth (Lawless & Heymann, 2013). Samples were placed in ice-filled preservation boxes. Data for the evaluation results were collected through the sensewhisper online sensory system (https://www.sensewhipser.com).

2.7. Image-recognition experiment configuration

2.7.1. Precise and efficient fish image detection using YOLOv5s

YOLO (You Only Look Once) is an object detection algorithm that follows a single-stage approach, treating the detection task as a regression problem (Redmon et al., 2016; Tian et al., 2019; Wang & He, 2021). It predicts the bounding box coordinates and class probabilities for the target object. YOLOv5 represents the most recent iteration in the YOLO (You Only Look Once) series (Wang & He, 2021; Zhu et al., 2021). YOLOv5 incorporates the utilization of CSPNet (Cross Stage Partial Network) and PANet (Path Aggregation Network) (Chen et al., 2021), with CSPNet serving as the backbone network for extracting image features, distinguishing it from earlier versions. It effectively balances model inference speed and accuracy while reducing the overall model size (Wang et al., 2020). PANet (Path Aggregation Network) functions as the neck network, facilitating improved feature fusion (Liu et al., 2018). Moreover, the open-source release of YOLOv5′s source code has gained popularity among researchers and developers. YOLOv5s represents a compact variant within the YOLOv5 series, denoting the “small” version of YOLOv5 (Song et al., 2021; Tang et al., 2021). YOLOv5s stands out as a lightweight model compared to its counterparts (Tang et al., 2021; Xue et al., 2022). Consequently, in this study, we employed YOLOv5s, characterized by the smallest model size among the YOLOv5 models.

The network architecture of YOLOv5s is depicted in Fig. 2(a), comprising four main components: Input, Backbones, Neck, and Head. The fundamental components of the YOLOv5s-based detection network are as follows:

(1) Focus Module: This module slices the input image into smaller parts, expanding the input channels while preserving complete image information. It effectively extracts features using Depth-wise Separable Convolution.
(2) CBL Module (Convolution, Batch Normalization, and Leaky ReLU Activation): This module is responsible for feature extraction from the image (Ting et al., 2021). It utilizes convolution and batch normalization techniques, applying the Leaky ReLU activation function expressed as Eq. (1):

$$L(x) = \begin{cases} xifx \geq 0 \\ \lambda xifx < 0 \end{cases} \quad (1)$$

Note: Where λ is a number within the range (0, 1).

(3) Cross Stage Partial Network (CSPNet) Module: Two types of CSP structures are utilized in this network. The CSP1_X structure is implemented in the backbone network, incorporating CBS modules, multiple residual units, and a concatenation function to enhance the learning capability of the convolutional neural network. On the other hand, the CSP2_X structure is employed in the neck component, comprising a CBS module and a concatenation function to enhance the feature integration capability of the network (Ting et al., 2021).
(4) Spatial Pyramid Pooling-Fast (SPPF) Module: This module plays a crucial role in capturing information from objects of various sizes. By performing pooling operations (Maxpool) on features at different scales, the SPPF module enables the network to effectively handle objects at different spatial resolutions (Ting et al., 2021).

2.7.2. Enhanced fish image detection with improved YOLOv5s

The improved model incorporates a small object detection layer into the network structure’s neck and integrates the convolutional block attention module (CBAM) into the convolutional module to optimize the model. Fig. 2(b) illustrates the network architecture of improved YOLOv5s.

2.7.2.1. Incorporating a small object output layer. This study introduces an enhancement to the YOLOv5s model through the addition of a dedicated small object output layer, as depicted in Fig. 2(b) within the modification area. The purpose of this layer is to enhance the representation and classification of small-sized objects by collaborating with the existing outputs. An important aspect of the small object detection layer involved adjusting the size of the output feature map. The improved YOLOv5 model set the output feature map size of the new small object detection layer to 1/4 of the original image size. Consequently, this modification allowed the network to capture finer details and extract more local information pertinent to small objects, leading to an enhancement in detection accuracy (1.3%). Furthermore, this study introduced novel anchor settings, namely [5,6], [8,14], and [15,11], for small object detection. The determination of these anchor settings employed k-means clustering to optimize their alignment with the size and aspect ratio of small objects, thereby ensuring enhanced localization and classification accuracy.

2.7.2.2. Data augmentation. To enhance the model’s object detection capability, the copy-paste augmentation technique was employed for data augmentation. Copy-paste augmentation is commonly used in object detection tasks for this purpose (Ghiasi et al., 2021). In this study, a source image was initially selected, and two target images were randomly chosen from the training dataset. For each selected target image, a rectangular annotation box was defined. The center of the box remained unchanged while the width and height were multiplied by a random scaling factor ranging from 1.5 to 2.5. This scaling resulted in the creation of a region of interest (ROI) (Ghiasi et al., 2021). The ROI was then sliced and copied to a random position within the source image. During the copying process, care was taken to avoid any overlap with existing positive samples in the source image. The number of copies was randomly determined, allowing the ROI of each target image to be copied 0 to 3 times. These operations increased the final number of positive samples in the source image, ranging from 1 to 7. Importantly, the added positive samples preserved the original background distribution. As a result, this data augmentation method significantly expanded the size of the dataset while maintaining the integrity of the original data distribution, thereby facilitating model updates (Ghiasi et al., 2021).

2.7.2.3. Incorporating the channel and spatial attention mechanism (CBAM) module. In this study, the Channel and Spatial Attention

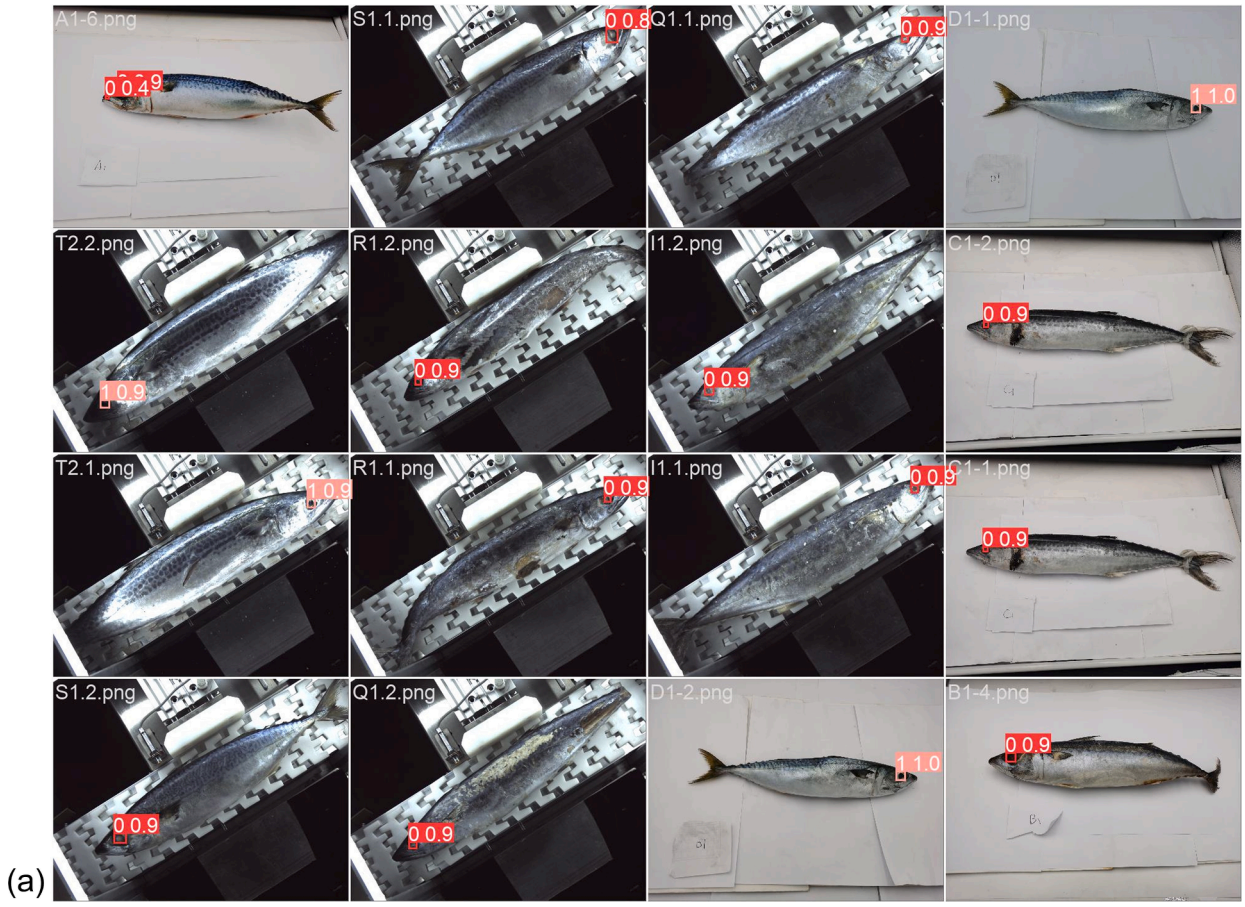


Fig. 3. Mackerel picture recognition results. *Note.* (a) Detection results of YOLOv5s; (b) Label Infographic; (c) PR curve changes of YOLOv5s; (d) F1-confidence changes of YOLOv5s; (e) Detection results of improved algorithm; (f) PR curve changes of improved algorithm; (g) Changes of mAP@0.5 during training of two algorithms; (h) Changes of mAP@0.5:0.95 during training of two algorithms; (i) Loss curve of bounding box regression during training; (j) Loss curve of classification during training; (k) Loss curve of bounding box regression during validation; (l) Loss curve of classification during training validation. Origin-YOLOv5s, Improve-improved YOLOv5s.

Mechanism (CBAM) module is added to the Convolution model in the head section. CBAM is an attention mechanism module for Convolutional Neural Networks (CNNs) (Xue et al., 2021). It was initially proposed in 2018 to enhance the performance of CNNs by leveraging attention on both channel and spatial dimensions. The CBAM module consists of two sub-modules: Channel Attention Module (CAM) and Spatial Attention Module (SAM). The CAM enhances useful features of the input feature map by learning weighted averages in the channel dimension and reduces the influence of irrelevant features. The SAM focuses on the importance of each pixel in the feature map and assigns higher weights accordingly.

(1) CAM module generates two (C, 1, 1) features based on spatial MaxPool and AvgPool operations, which are then input to a Multilayer Perceptron (MLP) to produce two new (C, 1, 1) features. These two features are element-wise summed and passed through a Sigmoid activation function to generate channel attention weights (M_c). Finally, the M_c is multiplied element-wise with the input feature F to obtain the final feature F' after the channel attention module.
(2) SAM module is also an attention mechanism used to enhance spatial information in the CNN. This module operates on the output feature F' from the channel attention module, with dimensions (C, H, W). It generates two (1, H, W) features based on channel-based MaxPool and AvgPool operations. These two features are concatenated along the channel dimension to obtain a (2, H, W) feature. Then, a 7x7 convolution operation is applied to reduce the channel dimension to 1, resulting in a (1, H, W) feature. This feature is then passed through a Sigmoid activation function to generate spatial attention weights (M_s). Finally, the M_s is multiplied element-wise with the feature F' to obtain the final feature after the spatial attention module. This improvement enables the model to better handle objects with extreme aspect ratios and enhances the accuracy of detection and classification.

2.7.3. Experiment configuration

The Ubuntul 8.04 system was used as the platform, Anaconda software was used for version control of the libraries used, and Pytorch l.7.0 was used as the training framework. CUDA 10.2 was chosen as the computational tool for floating-point operations (floating-point operations per second). The experimental code was run on an NVIDIA GTX 3090 with 24G of video memory; the use of this graphics processing unit significantly reduced the training and testing time. During the training, the optimizer used was Adam (momentum = 0.9, decay = $1e^{-5}$), the batch_size was set to 16, and the initial learning rate was set to 0.01. During the training process, the first three epochs were used as a warmup and the cosine increased to 0.01. After that, each epoch linearly decayed to 95%, and training was performed for a total of 50 epochs.

2.8. Data analysis

SPSS software (SPSS Inc., Chicago, IL, USA) was used to statistically

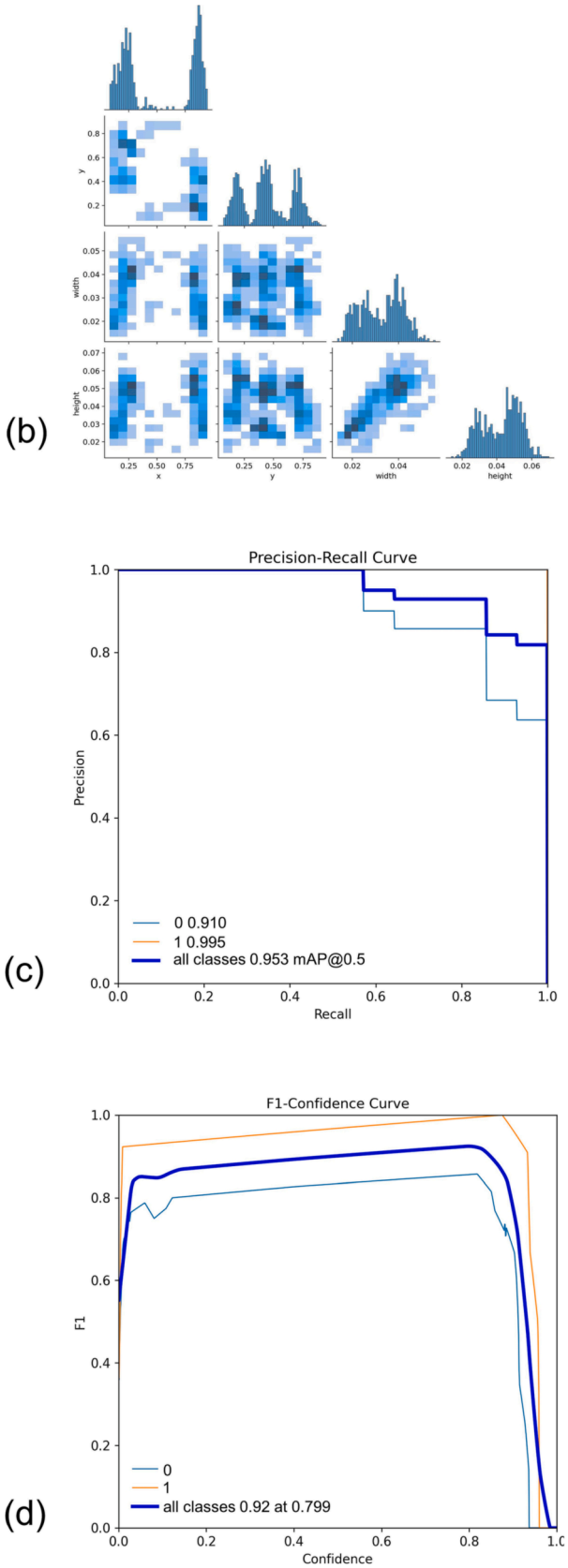
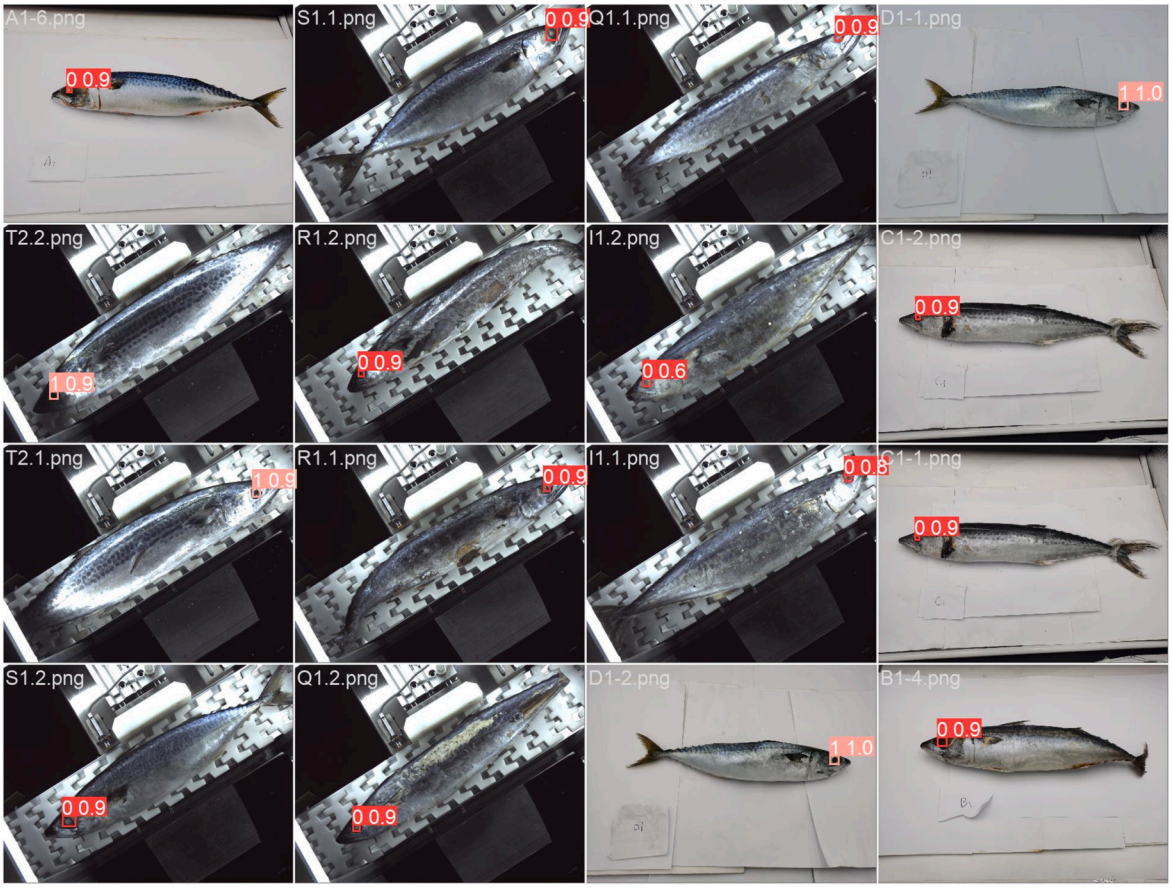
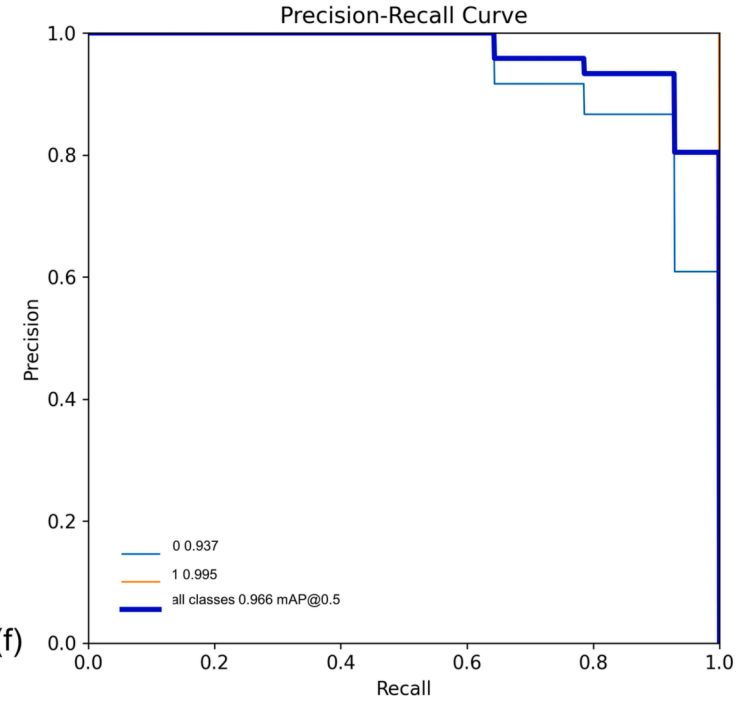


Fig. 3. (*continued*).

analyze the experimental data by one-way analysis of variance (ANOVA) and correlation analysis (Pearson) with $p < 0.05$ representing significance. The correlation map was drawn on ChiPlot (https://www.chiplot.online/). Origin 8.5 software (OriginLab Corporation, MA, USA) was used to plot the data results.



(e)



(f)

Fig. 3. (*continued*).

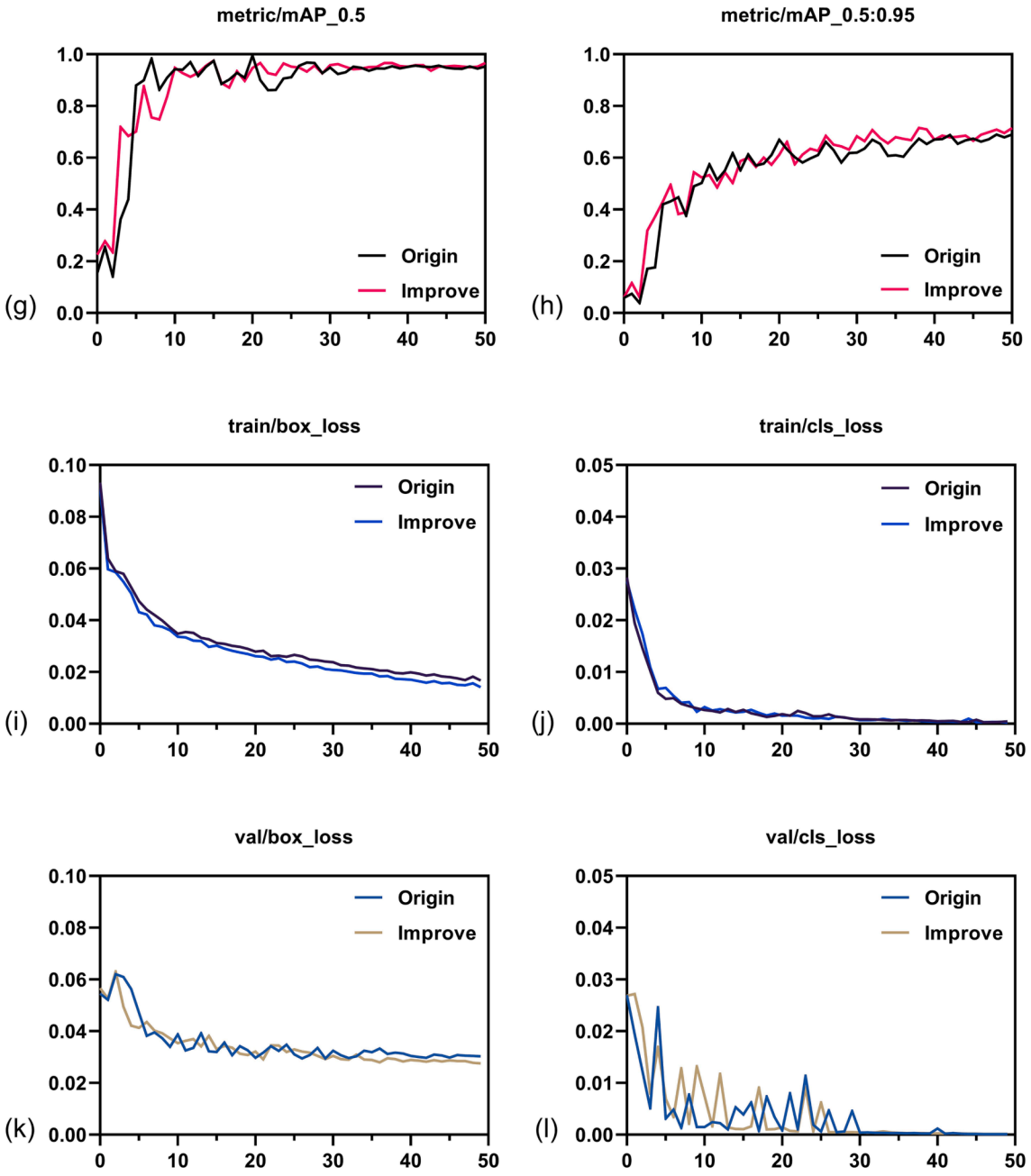


Fig. 3. (*continued*).

3. Results and discussion

3.1. Mackerel-picture recognition results

A total of 1100 sets of appearance feature images of Pacific chub mackerel and Spanish mackerel from 2019 to 2021 were collected to construct the dataset. A fast recognition model based on appearance–quality association was established by deep learning, and the results after loading and running are shown in Fig. 3(a). The letter + number.png is the number of the image corresponding to the fish sample. There are two sets of red numbers displayed at the fish-eye position, where 1 and 0 indicate whether the fish was from the current-year batch, and the number between 0 and 1 after that indicates the confidence rate. If the fish sample was tested as the current-year batch (2021), a 1 is displayed, meaning the sample passed. If the fish sample was tested as a non-current-year batch (2019, 2020), 0 is shown, indicating that the sample failed. The computing speed was 150 ms/image with the NVIDIA GTX 3090. Fig. 3(b) shows the label information plot of the data training set, where x represents the horizontal coordinate of the label centroid, y the vertical coordinate of the label centroid, width indicates the width of the label, and height is the height of the label. It can be seen by (x, y) that the label localization positions are mostly distributed on both sides of the image; the fisheye localization was relatively small compared with the whole image, as shown in Fig. 3(b); and the samples were mostly distributed in two regions: (0.02, 0.03), (0.04, 0.05–0.06).

PR curves were plotted by combining the results of accuracy and recall. As shown in Fig. 3(c), the higher the accuracy, the lower the recall. The area under this curve is called the AP value (average precision), and higher AP values represent better performance. The mAP (average precision) is the average of the AP values for each class and is used to represent the performance of multi-class label prediction. When the mAP is higher, the performance is better. In this model, the accuracy

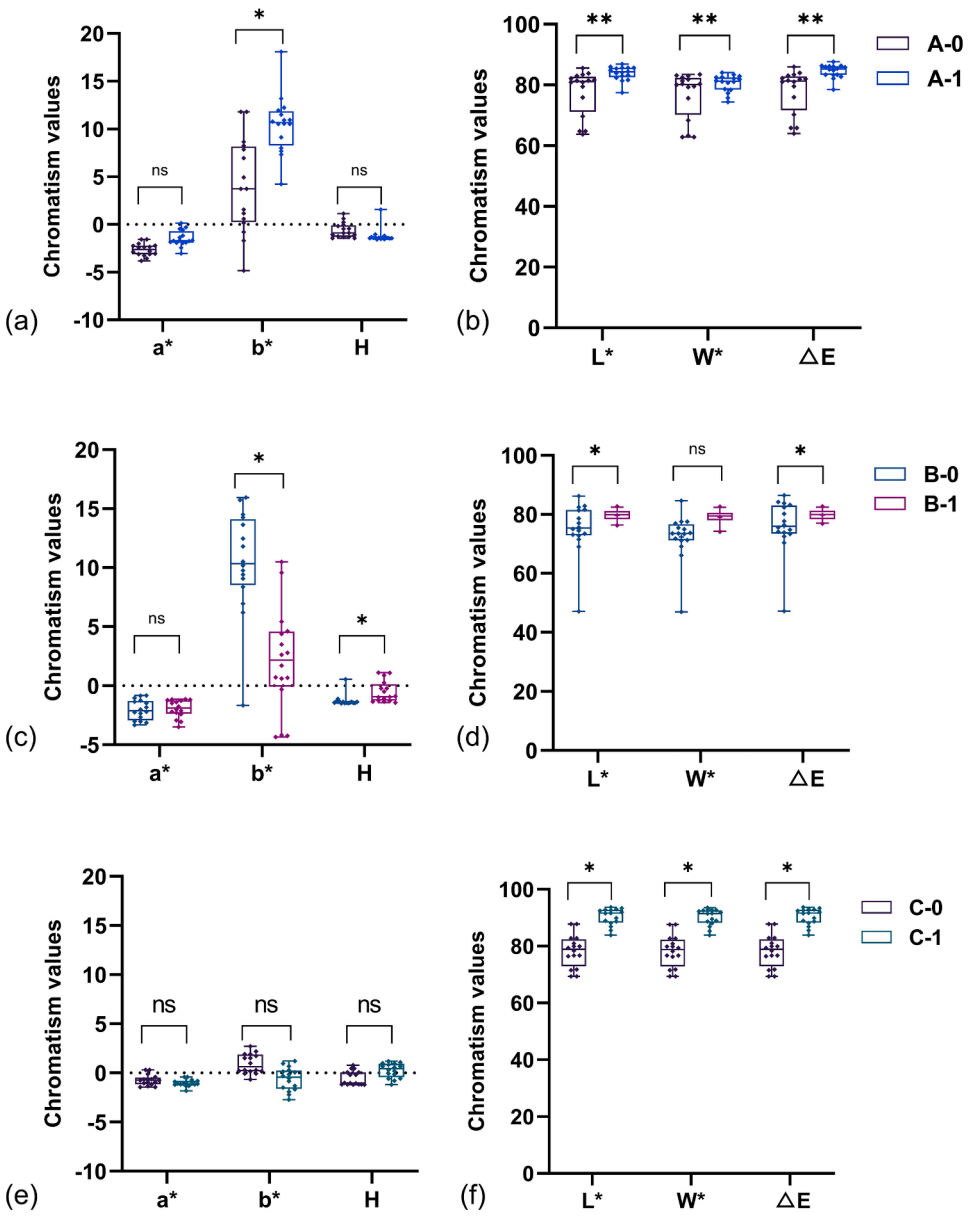


Fig. 4. Chromatic aberration results. *Note.* (a-b) Chromatism value of Group I samples; (c-d) Chromatism value of group II samples; (e–f) Chromatism value of Group III samples. (A) Group I samples, Pacific chub mackerel obtained from China; (B) Group II samples, Pacific chub mackerel obtained from Japan; (C) Group III samples, Spanish mackerel obtained from China. 0 indicates that the sample is not the current year batch in the initial sieving of image recognition, and the sample is not passed. 1 indicates that the sample is the current year batch in the initial sieving of image recognition, and the sample is passed. * indicates significant difference between samples ($p < 0.05$), ** indicates significant difference between samples ($p < 0.01$), and ns indicates no significant difference between samples.

at convergence, mAP@0.5, was 0.953, indicating that this method is able to detect objects stably. The F1 score is the average of precision and recall (0–1, where 1 represents the best and 0 represents the worst) and is a measure of the classification problem (Fig. 3(d)). In this model, the F1 score reached 0.92 in multiple categories. It is worth noting that there are a few samples with unsatisfactory detection results shown in Fig. 3 (a), such as A1-6, where the confidence of the judgment results was 0.4. For this reason, the model performance was further optimized in this study.

3.2. Evaluation metrics of the algorithm after improving the model

Considering that YOLOv5 is very suitable for applications in embedded devices because of its fast detection speed and relatively high recognition accuracy. However, it is not applicable to items with extreme aspect ratios (which would result in low-confidence predictions). Therefore, the k-means clustering algorithm was used to set the initial anchor box, increase the anchor-box dimensions, and perform model optimization. In Fig. 3(e), the prediction results of the improved model are represented. The improved model has a confidence of 0.9 when performing mackerel judgments of numbers A1–6. To verify the performance of different methods and models, PR curve plotting was performed. In Fig. 3(f), the PR curve of the improved model is depicted. The higher the precision, the lower the recall rate. The mAP@0.5 value of the improved model in this study is 0.966. Compared with the original, the improved algorithm fits the total class better with a higher mAP@0.5 value and better performance. To represent the variation in the mAP metric, we recorded the variation in mAP@0.5 and mAP@0.5:0.95 in our experiments and visualized it in graphical form (Fig. 3(g-h)). The results showed that the metrics of improved model is higher than the metrics of original model, and, therefore, the tuning of the original model resulted in improved performance. In addition, this study compared the loss curves of the models before and after the improvement, including bounding-box regression in training (Fig. 3(i)), classification loss in training (Fig. 3(j)), bounding-box regression in validation (Fig. 3(k)), and classification loss in validation (Fig. 3(l)). The train/box_loss indicated the error (GIoU) between the prediction and calibration boxes in the training set. The smaller the average value of the GIoU loss function, the more accurate the localization. The train/cls_loss indicated the correctness of the calculated anchor box and the

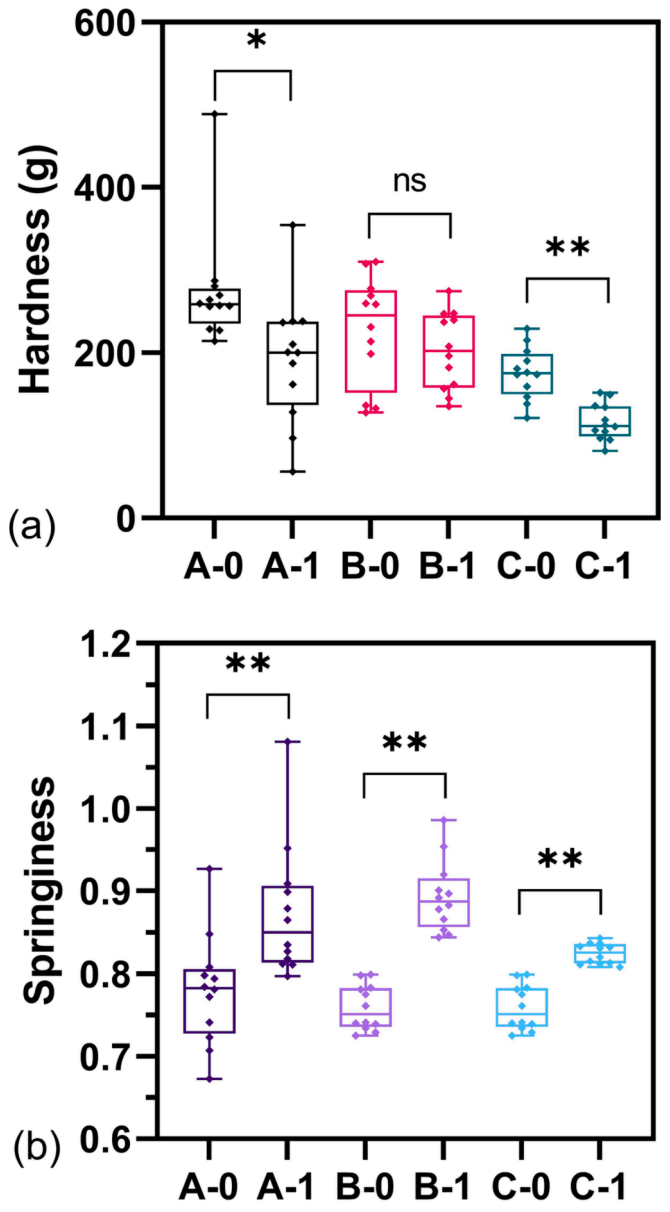


Fig. 5. Texture analysis results. *Note.* (a) Hardness analysis of samples; (b) Springiness analysis of samples. (A) Group I samples, Pacific chub mackerel obtained from China; (B) Group II samples, Pacific chub mackerel obtained from Japan; (C) Group III samples, Spanish mackerel obtained from China. 0 indicates that the sample is not the current year batch in the initial sieving of image recognition, and the sample is not passed. 1 indicates that the sample is the current year batch in the initial sieving of image recognition, and the sample is passed. * indicates significant difference between samples ($p < 0.05$), ** indicates significant difference between samples ($p < 0.01$), and ns indicates no significant difference between samples.

corresponding calibration classification, and the smaller the average value of the classification result, the more accurate the classification. The val/box_loss and val/cls_loss represented the localization loss and classification loss of the validation set, respectively. The results showed that the validation set converged at 30 epochs. Comparing the two models, the edge loss and classification loss were almost identical for both models, indicating that the adjustment to the model had little effect on the loss of bounding-box regression and classification. Therefore, in this study, the improved-model screening results were used for the next detection steps.

3.3. Colorimeter characterization

Since the freshness of whole fish is generally judged by the color of the fish eyes, gills, and abdomen, the location of the abdomen, 2 cm after the gills and 1 cm below the fins, was selected for color difference detection in this study. The color of the fish abdomen was described through six characteristics of the colorimeter, namely, L* (brightness), a* (redness/greenness), b* (yellowness/blueness), W (whiteness), ΔE (color difference value), and H (hue). In this study, color difference analysis was performed for the current-year batch samples (shown as passed samples, 1) and non-current-year samples (shown as failed samples, 0) identified by the model. Considering the differences in the fish samples, three groups of samples, Pacific chub mackerel from China (Group I, A), Pacific chub mackerel from Japan (Group II, B) and Spanish mackerel from China, were tested separately. There were 24 samples in each group (12 samples passed the initial screening and 12 samples failed the initial screening).

As shown in Fig. 4(a-b), in the first group of samples, the L*, W*, ΔE, and b* values of the samples that passed the picture test were significantly higher than those of the failed samples. In the second group of samples (Fig. 4(c-d)), the passed samples were significantly higher in the L*, ΔE, b*, and H values than in the failed samples. In addition, the third group of samples included significant differences in the L*, W*, and ΔE values (Fig. 4(e-f)). In summary, the color difference values L* and ΔE in the qualified samples in all three groups were significantly higher than those of the unqualified samples. One possible reason for this is that, with the extension of freezing time, the color of the abdomen of the samples gradually darkened and the brightness, whiteness, and color difference gradually decreased. However, the b* values of the qualified samples in group 1 were higher than those in the failed group, and the b* values in the qualified samples in group 2 were lower than those in the failed group due to the individual differences between the samples.

It has been shown that chromatograph disruption during fish freezing causes yellowing of the fish, with subsequent release and spreading to the subcutaneous layer, as well as brown discoloration due to lipid oxidation. As a result, mackerel skin tends to lose its color with longer storage times, changing from iridescent blue to light blue (dorsal) and from pearl/white to gold (abdomen) (Alfama et al., 2009). The image established that the freezing time of the samples in the qualified group was shorter than that of the unqualified group; thus, the change in color between the two groups was in accordance with the color changes that generally occur during frozen mackerel storage.

3.2. Texture analysis

Textural analysis of the primary sieved samples was carried out in this study. Fig. 5(a) shows the hardness of the three groups of samples, and Fig. 5(b) shows the springiness of the three groups of samples. The results of the qualitative analysis indicated that there was a significant difference in hardness and springiness between the passed and failed samples in the first group of samples. The springiness of the passed samples in the second group was significantly different from that of the failed samples. The passed samples in the third group were significantly different from the failed samples in terms of hardness and springiness. One possible reason for this is that, with a longer frozen time, the drip loss from the fish muscle increases, in addition to the increased hardness and decreased springiness of fish flesh caused by cell damage and protein aggregation (Cropotova et al., 2019). The images established that the freezing time of the samples in the passed group was shorter than that of the failed group; thus, the texture changes in both groups were consistent with the texture changes resulting from frozen mackerel storage.

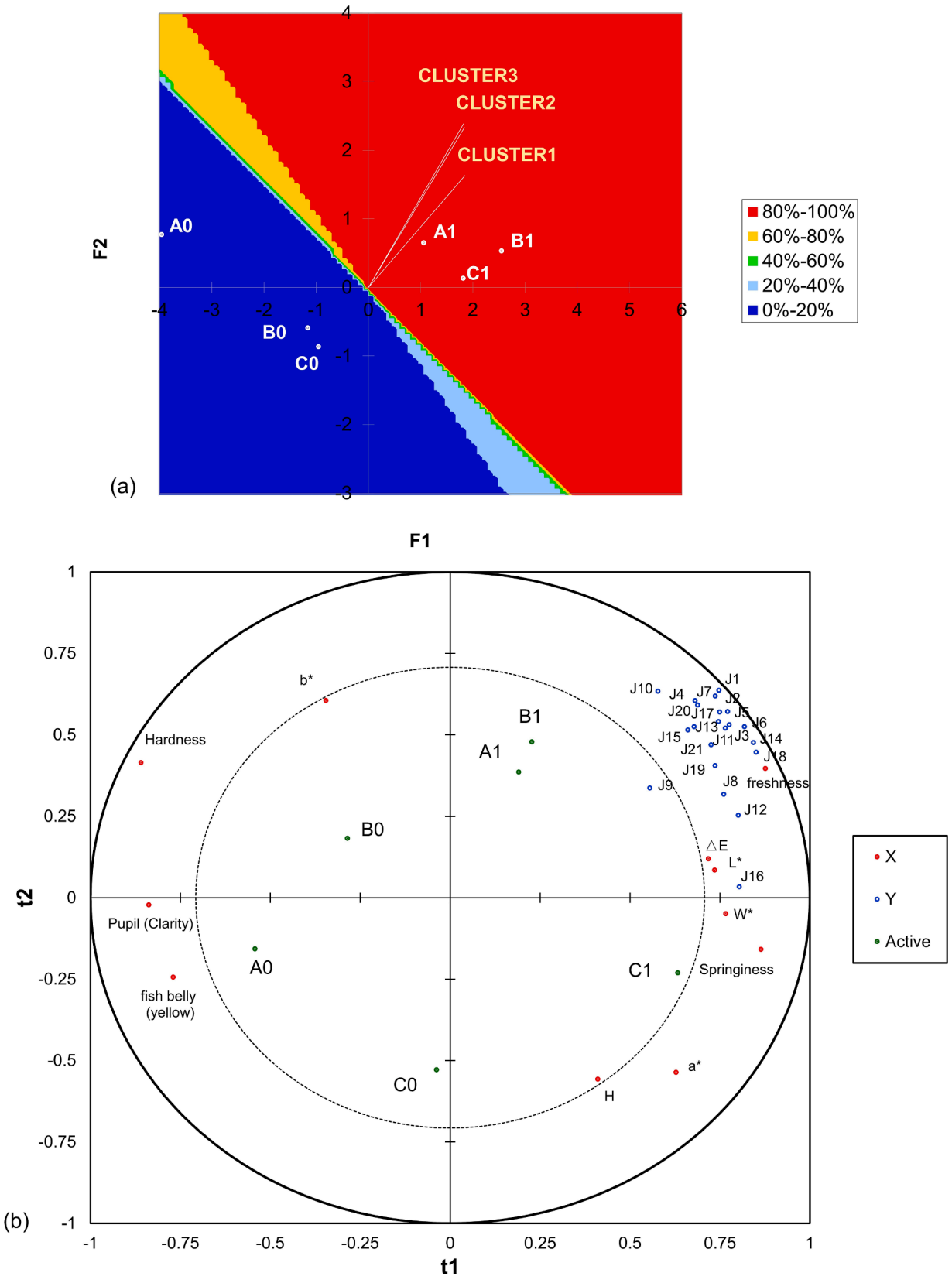


Fig. 6. Sensory analysis. *Note.* (a) Preference mapping of different samples; (b) PLSR analysis of sensory attribute and instrument parameters; (c) Correlation analysis of different parameters. (A) Group I samples, Pacific chub mackerel obtained from China; (B) Group II samples, Pacific chub mackerel obtained from Japan; (C) Group III samples, Spanish mackerel obtained from China. 0 indicates that the sample is not the current year batch in the initial sieving of image recognition, and the sample is not passed. 1 indicates that the sample is the current year batch in the initial sieving of image recognition, and the sample is passed.

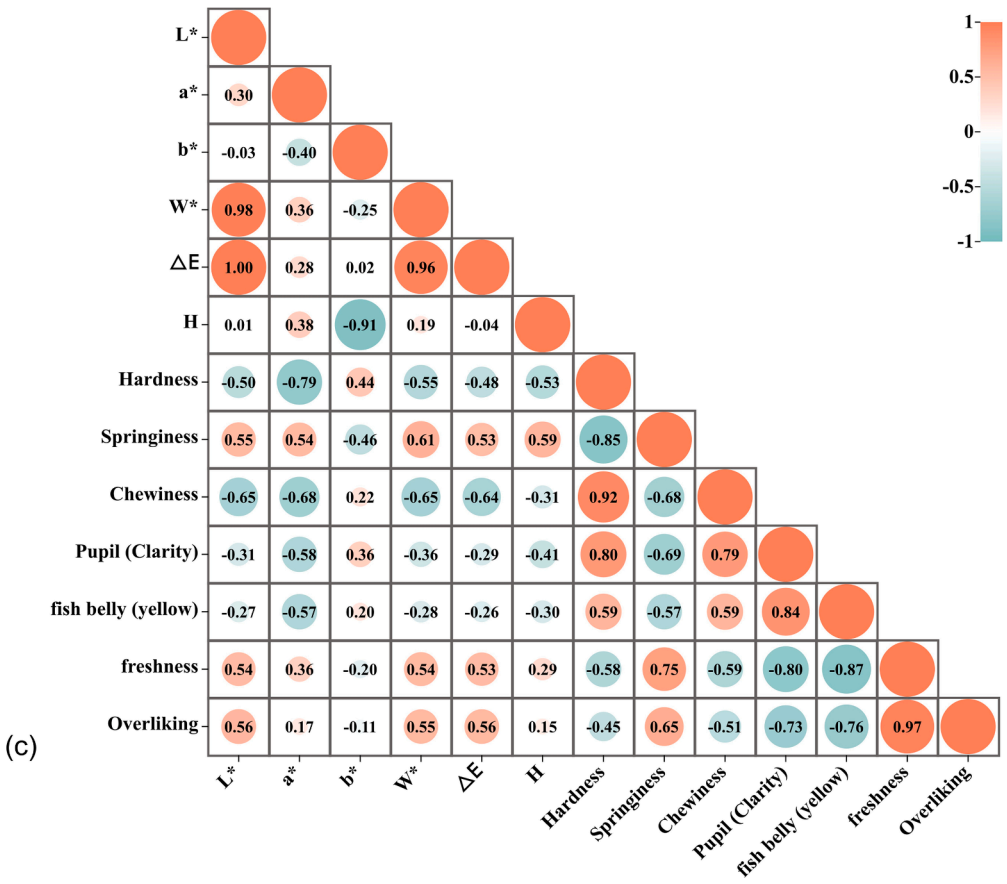


Fig. 6. (*continued*).

3.5. Sensory characterization and the relationship between quality characteristics

To better understand consumer attitudes, sensory evaluation was used to evaluate the samples. The preferences of the 21 assessors were divided into three clusters by cluster analysis (AHC). As displayed in the Fig. 6(a), the contour areas of different colors show whether the clusters have above-average preferences. The preference scores of the three passed samples ranged from 80% to 100%, as illustrated in Fig. 6(a). The preference scores of the failed samples in the three clusters range from 0% to 20%. This indicated that consumers could clearly distinguish between passed and failed samples and preferred the samples that passed.

With sensory attribute scores, color difference values, and quality construct values as independent variables, the explained *X* variable variance contributions were PC1 = 55% and PC2 = 25%. With overall evaluator preference as the dependent variable *Y*, the explained variance contribution of the *Y* variable was PC1 = 55% and PC2 = 12%. The algorithm considered extracting as many principal components as possible from *Y* and *X* (principal component analysis, PCA) when building the regression, and performing the extraction from *X* and *Y* separately (Fig. 6(b)). In this study, the PLSR model reached values of 0.935, 0.879 and 0.581 for R^2X(cum), R^2Y(cum), and Q^2(cum), respectively. This indicated that the components generated by PLSR summarize the *X* and *Y* cases well. As shown in the Fig. 6(b), the variables in the middle of the two circles were well-represented in both dimensions. The pupil (clarity) and fish belly (yellow) of the sample had a strong positive correlation with the hardness of the sample. The freshness, L*, W*, ΔE and springiness of the samples showed strong positive correlations with the evaluator preference scores. Moreover, the hardness of the samples had a significant negative correlation with springiness, freshness, and consumer preference scores.

The sensory results were further analyzed with physicochemical indicators and the Pearson values were obtained from Pearson correlation coefficient analysis (Fig. 6(c)). The correlation analysis was performed by combining the sensory scores with the colorimeter and texture results, and the results showed that the overall preference and freshness scores were significantly correlated. The scores of pupil and fish belly were significantly and negatively correlated with overall liking and freshness. This indicated that the cloudier the fish pupil and the more yellow the fish-belly color, the less fresh the fish was perceived to be by the evaluator and the lower the preference. In addition, springiness was significantly correlated with overall preference and freshness, suggesting that the more elasticity detected by the mass spectrometer, the more consumers tend to perceive the fish as fresher and prefer the sample. This finding was largely consistent with the PLSR analysis.

4. Conclusion

This study demonstrated the efficacy of a primary screening method for Pacific chub mackerel and Spanish mackerel samples based on the target detection model of YOLOv5s. The results of model accuracy and recall rate analysis indicated that the model was able to detect targets stably and that the performance of the improved model algorithm was better than that of the pre-improvement model. The model's accuracy was assessed through sensory evaluation, texture profile analysis, and colorimeter analysis. The findings indicated that the enhanced model achieved an mAP@0.5 score of 0.966, surpassing the original version's score of 0.953. Moreover, the improved model's params was only 7.848 M, the average detection time was 115 ms/image (image resolution 2400x3200, NVIDIA GTX 3090). Furthermore, sensory and physicochemical indicators reliably distinguished between qualified and

unqualified samples. The associated PLSR model exhibited R^2X, R^2Y, and Q^2 values of 0.977, 0.956, and 0.663, respectively. The PLSR analysis showed that a decrease in springiness, an increase in hardness, cloudy eyes, and the yellowing of the belly resulted in lower preference values from the assessors. These results were generally consistent with the findings of the correlation analysis. This study provides new insights into the rapid sorting and fast sensory characterization of traditional fish products through the inclusion of fast picture screening based on deep learning. Moreover, this study offers a revolutionary solution for the real-time monitoring and analysis of fish product quality. By allowing for the immediate evaluation of fish products, this innovation provides consumers with the assurance that they are purchasing high-quality seafood. To improve the usefulness of the model, further research is still needed.

Funding

This work was supported by the National Key R&D Program of China [2019YFD0902000]; and Key Science and Technology Program of Liaoning Province [2020JH1/10200001].

CRediT authorship contribution statement

Yi-Zhen Huang: Conceptualization, Data curation, Writing – review & editing, Writing – original draft. **Lin Han:** Methodology, Formal analysis, Writing – original draft. **Xiaoqing Yang:** Formal analysis, Data curation, Writing – review & editing. **Yu Liu:** Writing – review & editing. **Bei-Wei Zhu:** Supervision, Writing – review & editing. **Xiu-Ping Dong:** Funding acquisition, Supervision, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors would like to thank Youlian Li and Dong Zhang for their help with this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2023.100733.

References

Alfama, P. M., Sveinsdóttir, K., & Martinsdóttir, E. (2009). Quality Index Method (QIM) for frozen-thawed Atlantic Mackerel (Scomber scombrus) stored in ice: Development and Application in a shelf life study. Fisherries training programme. https://www.grocentre.is/static/gro/publication/217/document/patricia09prf.pdf.

Chen, Y., Zhang, C., Qiao, T., Xiong, J., & Liu, B. (2021). Ship detection in optical sensing images based on YOLOv5. In Twelfth International Conference on Graphics and Image Processing (ICGIP 2020) (Vol. 11720, pp. 102-106). SPIE.

Cropotova, J., Mozuraityte, R., Standal, I. B., Grøvlen, M. S., & Rustad, T. (2019). Superchilled, chilled and frozen storage of Atlantic mackerel (Scomber scombrus) fillets–changes in texture, drip loss, protein solubility and oxidation. International Journal of Food Science & Technology, 54(6), 2228-2235. https://doi.org/10.1111/ijfs.14136.

EUMOFA. (2019). European market observatory for fisheries and aquaculture products. 2019 Edition: The EU fish market. https://doi.org/10.2771/168390.

Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., & Zoph, B. (2012, 2021). Simple copy-paste is a strong data augmentation method for instance segmentation.

Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., & Zoph, B. (2021). Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2918–2928).

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Hyldig, G., Nielsen, J., Jacobsen, C., & Nielsen, H. H. (Eds.). (2012). Sensory and quality properties of packaged seafood. Advances in Meat, Poultry and Seafood Packaging, 154-170. . https://doi.org/10.1533/9780857095718.1.154.

ISO. (2012). Sensory analysis—General guidance for the selection, training, and monitoring of selected assessors and expert sensory assessors. *ISO Standard 8586*, 2012.

Lawless, H. T., & Heymann, H. (2013). *Sensory evaluation of food: Principles and practices*. Springer Science & Business Media.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Liu, L., Lan, W., Pu, T., Zhou, Y., & Xie, J. (2021). Combining slightly acidic electrolyzed water and slurry ice to prolong the shelf-life of mackerel (Pneumatophorus japonicus). *Journal of Food Processing and Preservation, 45*(9). https://doi.org/10.1111/jfpp.15762

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8759-8768).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.

Sanghyun, W., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

Song, Q., Li, S., Bai, Q., Yang, J., Zhang, X., Li, Z., & Duan, Z. (2021). Object detection method for grasping robot based on improved YOLOv5. *Micromachines, 12*(11), 1273.

Tang, J., Liu, S., Zheng, B., Zhang, J., Wang, B., & Yang, M. (2021). Smoking behavior detection based on improved YOLOv5s algorithm. In 2021 9th International Symposium on Next Generation Electronics (ISNE) (pp. 1-4). IEEE.

Thiansilakul, Y., Benjakul, S., & Shahidi, F. (2007). Compositions, functional properties and antioxidative activity of protein hydrolysates prepared from round scad (Decapterus maruadsi). *Food Chemistry, 103*(4), 1385–1394. https://doi.org/10.1016/j.foodchem.2006.10.055

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture, 157*, 417–426. https://doi.org/10.1016/j.compag.2019.01.012

Ting, L., Baijun, Z., Yongsheng, Z., & Shun, Y. (2021). Ship detection algorithm based on improved YOLO V5. In 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE) (pp. 483-487). IEEE.

Wang, C.-Y., Liao, H.-Y.-M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 390-391).

Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering, 210*, 271–281. https://doi.org/10.1016/j.biosystemseng.2021.08.015

Xue, J., Cheng, F., Li, Y., Song, Y., & Mao, T. (2022). Detection of farmland obstacles based on an Improved YOLOv5s algorithm by using CIoU and anchor box scale clustering. *Sensors, 22*(5), 1790. https://doi.org/10.3390/s22051790

Xue, M., Chen, M., Peng, D., Guo, Y., & Chen, H. (2021). One spatio-temporal sharpening attention mechanism for light-weight YOLO models based on sharpening spatial attention. *Sensors, 21*(23), 7949. https://doi.org/10.3390/s21237949

Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y. (2019). Application of deep learning in food: A review. *Comprehensive Reviews in Food Science and Food Safety, 18*(6), 1793–1811. https://doi.org/10.1111/1541-4337.12492

Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2778–2788).