

International Journal of Simulation and Process Modelling

ISSN online: 1740-2131 - ISSN print: 1740-2123

<https://www.inderscience.com/ijspm>

Defect detection method of Chinese mitten crab based on improved EfficientViT

Guangyu Mu, Heng Zhang, Zhongxiang Fei, Guodong Gao, Haiguang Zhang

DOI: [10.1504/IJSPM.2025.10071192](https://doi.org/10.1504/IJSPM.2025.10071192)

Article History:

Received:	10 February 2025
Last revised:	17 March 2025
Accepted:	02 April 2025
Published online:	01 September 2025

Defect detection method of Chinese mitten crab based on improved EfficientVit

Guangyu Mu, Heng Zhang and Zhongxiang Fei

Mechanical and Power Engineering College,
Dalian Ocean University,
Dalian 116023, China
Email: mugy1982@126.com
Email: 1276203587@qq.com
Email: 2156878951@qq.com

Guodong Gao*

Navigation and Shipbuilding Engineering College,
Dalian Ocean University,
Dalian 116023, China
Email: dlouggd@163.com
*Corresponding author

Haiguang Zhang

Mechanical and Power Engineering College,
Dalian Ocean University,
Dalian 116023, China
Email: 3161508037@qq.com

Abstract: An improved EfficientVit-based defect detection method for Chinese mitten crabs (*Eriocheir sinensis*) is proposed to enhance detection efficiency and accuracy while reducing human interference. A dataset of intact and defective crab images was constructed to improve the model's adaptability. The squeeze-and-excitation (SE) module and mix of experts hidden (MoH) module were integrated into the EfficientVit model to strengthen its feature extraction ability and sensitivity to subtle differences. The ReLU activation function was replaced with the parametric rectified linear unit (PReLU) activation function to address the 'dying neuron' issue and improve the model's sensitivity to negative inputs. The improved model achieved an accuracy of 97.34%, precision of 97.35%, recall of 97.34%, and F1-score of 97.34%, showing significant improvements over the baseline model. Future work will focus on optimising model performance and exploring its adaptability and generalisation in different environments.

Keywords: Chinese mitten crab; EfficientVit; defect detection; attention mechanism; activation function.

Reference to this paper should be made as follows: Mu, G., Zhang, H., Fei, Z., Gao, G. and Zhang, H. (2025) 'Defect detection method of Chinese mitten crab based on improved EfficientVit', *Int. J. Simulation and Process Modelling*, Vol. 22, Nos. 1/2, pp.47–59.

Biographical notes: Guangyu Mu received her Bachelor's degree in Thermal Energy and Power Engineering from Dalian Ocean University (DLOU), Dalian, China in 2003 and Master's degree in Mechanical Engineering from DLOU in 2007. She received her Doctor's degree in Mechanical Engineering from Dalian Jiao tong University, China in 2019. She is currently an Associate Professor at DLOU. Her research interests include the optimisation design and intelligent control of mechanical products.

Heng Zhang received his Bachelor's degree in Automotive Service Engineering from Shandong Huayu University of Technology, Dezhou, China. He is currently working toward his MS degree in Agricultural Engineering and Information Technology from DLOU. His study direction is the development of deep learning-based systems for gender recognition and automated sorting devices in Chinese mitten crabs (*Eriocheir sinensis*).

Zhongxiang Fei received his Bachelor's degree in Automotive Service Engineering from Shandong Huayu University of Technology, Dezhou, China. He is currently working toward his MS degree in Agricultural Engineering and Information Technology from DLOU, China. His research direction is the application of machine vision in industry.

Guodong Gao received his Bachelor's degree in Marine Engineering from DLOU, Dalian, China in 2003 and Master's degree in Mechanical Engineering from DLOU in 2007. He is currently an Associate Professor at DLOU, with main research focus on intelligent recognition technology for aquatic products.

Haiguang Zhang is currently pursuing his Bachelor's degree in Agricultural Intelligent Equipment Engineering at DLOU. His main research focuses on ocean intelligent equipment technology.

1 Introduction

Eriocheir sinensis, commonly known as the Chinese mitten crab, river crab or hairy crab, is one of the most important aquaculture species in China, highly valued for its delicious meat and high nutritional value (Xue et al., 2022; Luo et al., 2019). According to the China Fishery Statistical Yearbook 2022, the annual production of Chinese mitten crabs exceeds 700,000 tons, with a breeding area of over 666,700 hectares. As the aquaculture industry grows rapidly, the market demand for Chinese mitten crabs is increasing, and their quality directly affects consumer health and the economic benefits of breeding enterprises (Bai et al., 2023; Che et al., 2022). In Chinese mitten crab aquaculture, appendage defects – including missing claws, fractured legs, or incomplete pereopods – significantly impact commercial value and survival rates. Traditional manual inspection struggles with consistency due to morphological variability and labour-intensive workflows (Zhu et al., 2016). While machine vision-based detection equipment has been introduced to address these challenges, existing devices often exhibit low recognition accuracy and poor adaptability, particularly under complex backgrounds and variable lighting conditions (Zhang et al., 2024). Therefore, effective defect detection of Chinese mitten crabs is crucial for ensuring product quality and enhancing market competitiveness.

Recent advancements in deep learning have shown promise in crustacean quality assessment. Research on *Eriocheir sinensis* within this domain has primarily focused on two areas: object detection for biomass management and gender classification for selective breeding. In object detection, Cao et al. (2020) addressed real-time monitoring challenges by redesigning the single-shot detector (SSD) framework. Their approach combined MobileNetV2 as a lightweight backbone with a multi-scale feature pyramid network, enabling efficient detection of live crabs in turbid aquatic environments while achieving 99.01% mAP and 98.94% F1-score. Zhao et al. (2019) used YOLOv3 to detect the location and quantity of crabs after preprocessing a single underwater live crab image, achieving a mAP of 86.42%, a precision rate of 96.65%, and a recall rate of 91.30%. Zhang et al. (2023) proposed a method for detecting and recognising the carapace of Chinese mitten crabs based on YOLOv5 and principal component analysis (PCA), with a recognition rate of 84.42% and a matching speed of 1.859 seconds.

In the field of gender classification, Chen et al. (2023) proposed a lightweight Chinese mitten crab detection and

gender classification method based on an improved YOLOv4, named GMNet-YOLOv4, achieving a precision of 96.75%, recall of 97.07%, and mean average precision (mAP) of 97.23% on the test set. Gu et al. (2024) introduced a Chinese mitten crab juvenile gender recognition algorithm based on an improved Faster R-CNN, called C-SwinFaster, with a mAP of 98.45%, an increase of 10.33 percentage points over the original model. Cui et al. (2020) presented a new method for Chinese mitten crab gender classification using a deep convolutional neural network (CNN). This method first enriched the dataset with four different data augmentation techniques, then addressed the gradient disappearance issue and shortened training time through batch normalisation, and introduced dropout technology to suppress overfitting. The classification accuracy reached 98.90%, which is 3 percentage points higher than two state-of-the-art methods.

Despite these advancements, appendage defect detection remains underexplored. Current lightweight vision models like EfficientViT, which employs cascaded group attention blocks with parameter sharing for computational efficiency (Liu et al., 2023), face limitations in fine-grained defect recognition. Specifically, its static attention head configuration and uniform channel weighting limit adaptability to heterogeneous defect patterns, while the default ReLU activation functions risk neuron saturation during prolonged training cycles. These architectural constraints hinder precise localisation of subtle appendage defects, which require dynamic feature recalibration and robust negative input sensitivity.

To address these challenges, an improved EfficientViT-based method for detecting defects in Chinese mitten crabs is proposed, which can classify defective and intact crabs. The method uses the EfficientViT-M5 variant as the base model. Integrating the mix of experts hidden (MoH) module into the model enhances its ability to learn different feature dimensions and effectively distinguish between defective and intact crabs. The squeeze-and-excitation (SE) module is also incorporated to strengthen the model's representation of the features of defective and intact crabs. Replacing all rectified linear unit (ReLU) activation functions in the model with parametric rectified linear unit (PReLU) activation functions solves the 'dying neuron' problem and improves the model's sensitivity to negative inputs. Ablation experiments were conducted on the proposed model, and it was compared with several other deep learning models to verify its effectiveness.

2 Background and methodology

2.1 Dataset collection and preprocessing

Experimental data were collected from a seafood market in Dalian, using smartphones to capture one-crab-per-image photographs of intact crabs (no appendage defects) and defective crabs (exhibiting missing claws, fractured legs, or incomplete pereiopods). To simulate real detection environments, images were acquired under diverse lighting conditions (indoor, natural, and dim lighting) to enhance model adaptability. The raw dataset contains 1,008 images (345 intact-class images, 663 defective-class images) with an original resolution of $3,072 \times 3,072$ pixels in JPG format. Representative examples include defective crabs with missing left second pereiopod and right first pereiopod [Figure 1(a)] and missing left second/fourth pereiopods, right fourth pereiopod, and right cheliped [Figure 1(b)], while intact crabs with fully preserved appendages are shown in Figures 1(c)–1(d). All images were resized to a uniform size of 224×224 pixels to meet the input requirements of the EfficientVit model.

Small sample data can easily lead to poor model generalisation and high overfitting risk. To enhance the model's generalisation, data augmentation methods were employed. The augmentation process included horizontal and vertical flipping, random scaling, translation, rotation, and shearing with a certain probability, as well as adding Gaussian and salt-and-pepper noise. Each image was augmented six folds to simulate various situations that might be encountered in actual detection environments, thereby enhancing the model's generalisation. Finally, the augmented images were divided into training, validation,

and testing sets in a ratio of 8:1:1. Table 1 details the image count distribution across subsets.

Figure 1 Representative images of Chinese mitten crabs (see online version for colours)

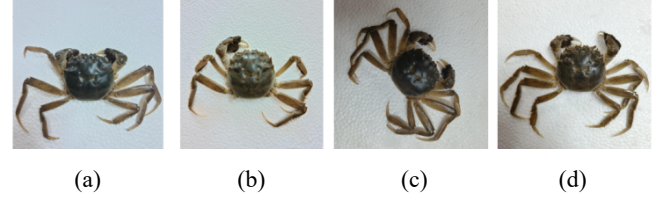


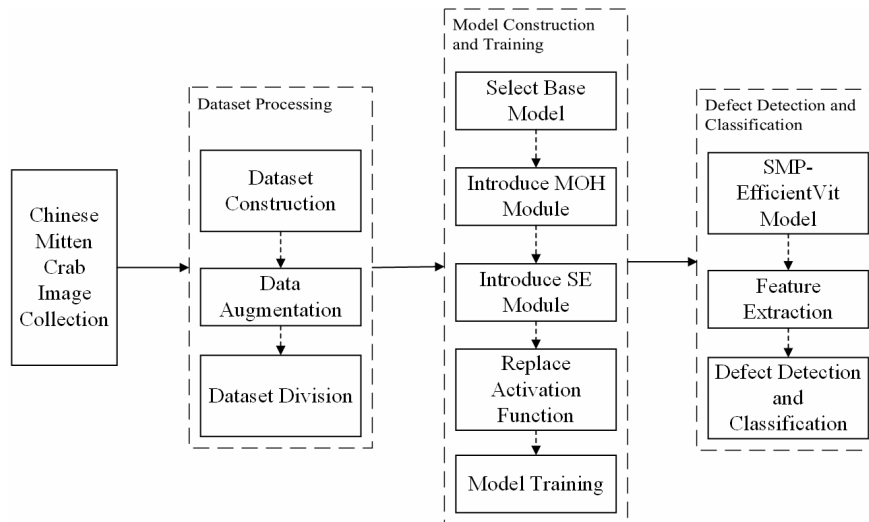
Table 1 Dataset partition (unit: images)

Dataset	Original quantity	Augmented quantity	Training set quantity	Validation set quantity	Testing set quantity
Intact crabs	345	2,415	1,932	238	245
Defective crabs	663	4,668	3,710	462	496

2.2 Overall workflow of the proposed method

An improved EfficientVit-based model for detecting limb defects in Chinese mitten crabs is proposed. The model, named SMP-EfficientVit, is based on the EfficientVit architecture and incorporates the SE module and MoH module to enhance feature extraction capabilities. Additionally, all ReLU activation functions in the model are replaced with PReLU activation functions. The overall process is shown in Figure 2.

Figure 2 Overall workflow of the proposed method



2.3 Model improvement plan

2.3.1 Overview of the EfficientVit base model

EfficientVit is a novel architecture designed to enhance the speed and memory efficiency of vision transformer models. It significantly accelerates the model's inference speed by optimising memory access and reducing computational redundancy (Liu et al., 2023). The architecture introduces a cascaded group attention module, which feeds different parts of the features into each attention head, reducing the computational redundancy of multi-head attention and enhancing feature learning capabilities. Additionally, it employs structured pruning techniques to reallocate parameters, improving parameter efficiency. This ensures that key network components have more channels to learn high-dimensional representations while reducing redundant parameters in less important modules, thereby accelerating the inference process and enhancing model efficiency. These improvements enable EfficientVit to maintain high accuracy while achieving faster inference speeds on different hardware platforms, making it particularly suitable for real-time applications and resource-constrained environments. The EfficientVit-M5 variant of the EfficientVit model is used as the base model.

2.3.2 Incorporating the mix of experts hidden module

The mix of experts hidden (MoH) module is an advanced attention mechanism that optimises the performance of multi-head self-attention (MHSA) by simulating the mixture-of-experts (MoE) framework (Jin et al., 2024). The core advantage of the MoH module is its ability to enable each input token to dynamically select the most relevant attention heads, rather than uniformly activating all heads. This design enhances the model's inference efficiency and introduces a weighted summation mechanism, thereby increasing the flexibility of the attention mechanism and unlocking additional performance potential.

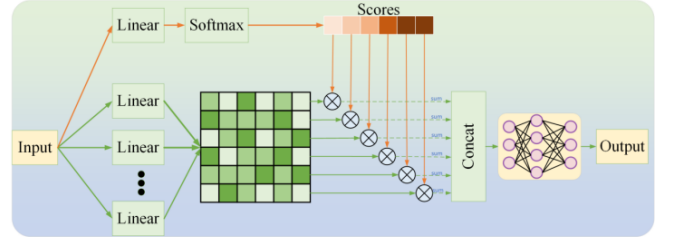
The MoH module consists of multiple attention heads and a router, which activates the top- K most relevant attention heads. The outputs of each attention head are weighted and summed to form the final output. The mathematical expression is shown in equation (1).

$$\text{MoH}(X, X') = \sum_{i=1}^h g_i H^i W_O^i \quad (1)$$

where g_i represents the routing score determining the activation weight of the i^{th} attention head, calculated through a learnable router network that implements top- K selection. The H^i term denotes the feature representation matrix generated by the i^{th} attention head through scaled dot-product attention between input tokens X (query) and X' (key/value), while W_O^i corresponds to the trainable output projection matrix specific to each attention head for dimension alignment and feature recombination. The summation index h indicates the total number of attention heads, serving as a hyperparameter that controls model

capacity. This design provides two key benefits: first, the routing mechanism enables dynamic head selection per token through sparse activation of relevant heads (where g_i becomes non-zero), improving inference efficiency while maintaining accuracy. Second, the weighted summation of transformed head outputs ($H^i W_O^i$) enhances attention flexibility compared to standard multi-head attention's uniform averaging, enabling better exploitation of heterogeneous feature patterns. The structure of the MoH module is shown in Figure 3.

Figure 3 Structure of the mix of experts hidden (MoH) module (see online version for colours)



The MoH module is integrated into the EfficientVit model's architecture to boost feature learning and representation. Within each EfficientVitBlock, the MoH module receives features from the previous layer through a linear layer (router) and dynamically assigns weights to different attention heads based on these features. Each linear attention head, also a linear layer, processes input features in parallel, outputting weighted feature representations. A softmax function normalises these outputs into probability distributions, generating scores that guide the attention mechanism. The MoH module then integrates these weighted features through another linear layer (output_proj), and a concat operation combines the outputs of multiple attention heads, producing the final feature representation fed into the next model stage. In this way, the MoH module improves the model's sensitivity to local features and enhances its ability to capture global features, improving the defect detection task.

2.3.3 Incorporating the squeeze-and-excitation module

The SE module is an innovative channel attention mechanism designed to enhance the feature representation capabilities of convolutional neural networks by explicitly modelling the interdependencies between channels (Hu et al., 2018). The module performs adaptive recalibration of feature maps through two main operations: 'squeeze' and 'excitation'. First, the SE module applies global average pooling to the feature map to generate a channel descriptor that captures the global information of channel features. Then, it uses two fully connected layers in the 'excitation' operation to learn the importance weights of each channel, which are used to recalibrate the channel responses in the feature map, thereby improving network performance. Specifically, the SE module includes three key steps: squeeze, excitation and reweight.

- 1 Squeeze operation: global average pooling is applied to the feature map, compressing the spatial dimensions (height and width) of each channel into a single value. The mathematical expression is shown in equation (2).

$$z = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (2)$$

where $x_c(i, j)$ represents the feature value at position (i, j) in the c^{th} channel, and H and W are the height and width of the feature map, respectively.

- 2 Excitation operation: two fully connected layers and a nonlinear activation function are used to learn the importance weights of each channel. The mathematical expression is shown in equation (3).

$$s = \sigma(g(z, W)) = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot z)) \quad (3)$$

where g denotes the fully connected layer, W_1 and W_2 are the weights of the fully connected layers, σ is the sigmoid activation function, and ReLU is the ReLU activation function.

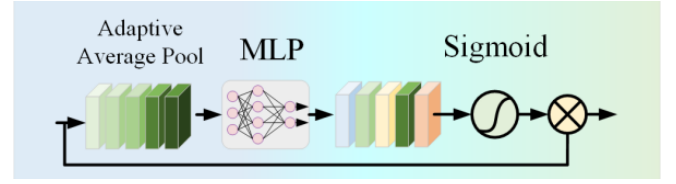
- 3 Reweight operation: the learned weights s is applied to each channel of the original feature map to achieve channel recalibration. The mathematical expression is shown in equation (4).

$$\hat{x}_c = s_c \cdot x_c \quad (4)$$

where $\hat{x}_c$ represents the recalibrated feature map, and s_c is the weight of the c^{th} channel.

The SE module is introduced to optimise the feature representation of the EfficientVit model. By appending the SE module after each EfficientVitBlock, the explicit modelling of the interdependencies between feature map channels is achieved, enhancing the model's ability to identify key features and thus improving the accuracy of Chinese mitten crab defect detection. The SE module consists of three main components: adaptive average pool, MLP and sigmoid. The adaptive average pool layer compresses the spatial dimensions of the feature map into a single value per channel. This is followed by the MLP, which typically includes two fully connected layers, to learn the importance weights of each channel. The sigmoid function then normalises these weights to a value between 0 and 1, which are used to recalibrate the channel responses. Through these squeeze and excitation operations, the SE module adaptively recalibrates channel responses, strengthening the model's ability to capture both local and global features. This provides an effective feature enhancement mechanism for the defect detection task of Chinese mitten crabs. The structure of the SE module is shown in Figure 4.

Figure 4 Structure of the mix of experts hidden (MoH) module (see online version for colours)



2.3.4 Parametric rectified linear unit activation function

The PReLU activation function is a variant of the ReLU activation function. It introduces a learnable parameter to allow a small non-zero output for negative values (He et al., 2015). This adaptive negative value processing ability enables PReLU to solve the 'dying neuron' problem that ReLU may encounter, where the output of ReLU is always zero for negative inputs, causing the gradient to vanish and preventing the network parameters from updating. The definition of the PReLU function is shown in equation (5), where a_i is a learnable parameter that is typically optimised through the backpropagation algorithm during training. This adaptive negative value processing mechanism endows PReLU with stronger expressive power and sensitivity to input data.

$$f(y_i) = \begin{cases} y_i, & \text{if } y_i > 0 \\ a_i y_i, & \text{if } y_i \leq 0 \end{cases} \quad (5)$$

where y_i represents the input to the nonlinear activation function f in channel i . a_i is used to control the slope of the function. By adjusting a_i , different channels can use different activation functions. Since a_i can be learned adaptively, the function definition in equation (5) can be transformed into equation (6).

$$f(y_i) = \max(0, y_i) + a_i \min(0, y_i) \quad (6)$$

To enhance the model's sensitivity to negative inputs and strengthen its feature expression, all ReLU activation functions in the model were replaced with PReLU. This replacement was implemented through a custom function that recursively traverses all modules of the model, identifying ReLU and replacing it with PReLU without affecting other parts of the model. With this improvement, the model can more effectively utilise negative input information during training, which is expected to enhance the model's generalisation ability and overall performance. The adaptive nature of PReLU allows the model to automatically adjust the processing strategy for positive and negative inputs, increasing the model's flexibility and robustness in diverse datasets and different tasks. Specifically, in the task of Chinese mitten crab defect detection, the introduction of PReLU helps the model capture more subtle feature differences, thereby improving the accuracy and reliability of defect detection.

2.4 Chinese mitten crab defect detection model based on SMP-EfficientVit

A Chinese mitten crab defect detection model based on the improved EfficientVit, named SMP-EfficientVit, is proposed. The model has undergone several key improvements on the original EfficientVit to enhance its feature extraction ability and generalisation performance. These improvements include the introduction of the SE module and the MoH module, as well as the replacement of all ReLU activation functions with PReLU. These enhancements work together to enable the SMP-EfficientVit model to more effectively capture key features and increase sensitivity to subtle differences when dealing with Chinese mitten crab defect detection tasks.

The SMP-EfficientVit model begins with the input layer, where the input image is divided into small patches, and features are extracted through the patch embedding layer. Subsequently, the model enters multiple stages, each containing an SMP-EfficientVit block. These blocks integrate the SE and MoH modules to enhance feature

representation. The SE module uses global average pooling and fully connected layers to adaptively recalibrate channel responses, while the MoH module employs multiple attention heads and a router to dynamically assign weights to different features. At the end of each stage, the spatial dimensions of the feature map are gradually reduced, and the number of channels is increased through patch merging and downsampling operations to capture more abstract features. Finally, after passing through the global pooling layer, the features are sent to the linear layer for classification, outputting a binary classification result for defective and intact crabs. The structure and working process of the SMP-EfficientVit model are shown in Figure 5.

Table 2 lists the detailed parameters of the SMP-EfficientVit model structure, including the input and output dimensions of each stage, the size of the convolutional kernel, the stride, and the number of heads in the multi-head self-attention. These parameters collectively define the model's architecture, ensuring its effective handling of the Chinese mitten crab defect detection task.

Figure 5 Structure of the SMP-EfficientVit model (see online version for colours)

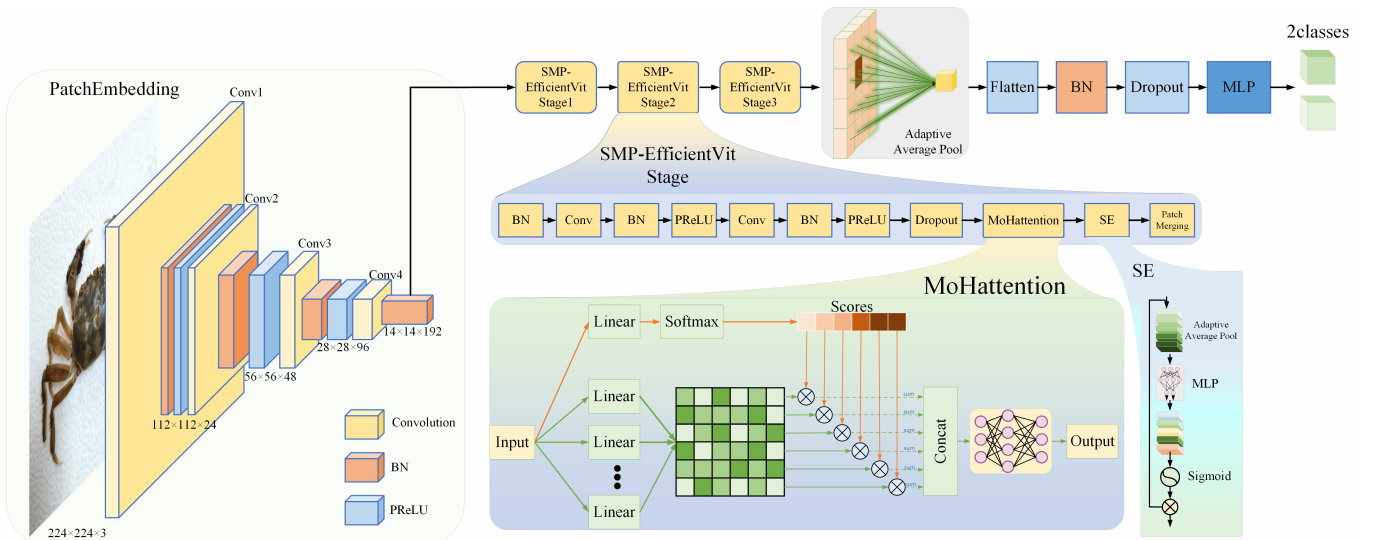


Table 2 Detailed parameters of the SMP-EfficientVit model structure

Stage	Component name	Input dimensions	Output dimensions	Kernel size	Stride	Number of heads	Remarks
Input	Input layer	3	192				
Patch embedding	Conv2d_BN	192	192	3	2		SE
Stage 1	SMP-EfficientVit block	192	192	3	1	8	SE+MoH
Patch merging	Conv2d_BN	192	288	2	2		Down sampling
Stage 2	SMP-EfficientVit block	288	288	3	1	8	SE+MoH
Patch merging	Conv2d_BN	288	384	2	2		Down sampling
Stage 3	SMP-EfficientVit block	384	384	3	1	8	SE+MoH
Global pooling	AdaptiveAvgPool2d	384	384				Global pooling
Head	Linear layer	384	2				Binary classification output

2.5 Model performance evaluation metrics

The performance of the defect classification model is evaluated using four complementary metrics – accuracy, precision, recall and F1-score. This multi-metric approach is essential because accuracy alone may fail to reflect model robustness in imbalanced datasets, such as ours, where the defective crab class is smaller than the intact class. Precision ensures the model minimises false positives, which is critical to avoid unnecessary economic losses from discarding intact crabs. Recall measures the model's ability to avoid false negatives, ensuring defective crabs are not misclassified as intact. The F1 score balances these trade-offs, providing a holistic assessment aligned with industrial requirements for automated quality control systems.

Accuracy (A), precision (P), recall (R), and F1-score (F1) are used to comprehensively evaluate model performance. Specifically, accuracy represents the proportion of correctly classified samples out of all samples; precision indicates the proportion of correctly classified samples within all samples predicted to be a certain class; recall represents the proportion of correctly classified samples within all actual samples of a certain class; the F1 score combines precision and recall to provide a comprehensive evaluation of the model, serving as the harmonic mean of precision and recall, and is used to measure the model's performance when precision and recall are balanced. The formulas for calculating each metric are as follows:

$$A = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (7)$$

$$P = \frac{TP_i}{TP_i + FP_i} \quad (8)$$

$$R = \frac{TP_i}{TP_i + FN_i} \quad (9)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (10)$$

where TP_i represents the number of correctly classified samples in class i ; FN_i represents the number of incorrectly

classified samples in class i ; FP_i represents the number of samples from other classes incorrectly classified as class i ; TN_i represents the number of samples from other classes not incorrectly classified as class i . Higher values of these four metrics indicate better model prediction performance, while lower values suggest poorer model prediction performance.

2.6 Experimental environment setup

The model training and testing were carried out under the Windows 11 operating system using a deep learning framework based on PyTorch, version 2.5.0. The hardware configuration included an 13th Gen Intel(R) Core(TM) i5-13500HX processor with a base frequency of 2.50 GHz, and 16 GB of DDR5 4,800 MHz memory. The graphics processing unit (GPU) was an NVIDIA GeForce RTX 4060 with 8 GB of video memory. The code compilation environment was PyCharm, with Python version 3.11.8. The construction and management of the experimental environment used the Conda environment manager, version 24.1.2.

3 Results and analysis

3.1 Comparison of training strategies

In the field of deep learning, a training strategy refers to a series of parameters and configurations determined before the start of model training. These strategies significantly impact the model's performance and its ability to learn image features. By carefully adjusting the training strategy, the model's learning process of image features can be optimised, thereby enhancing model performance and effectively preventing overfitting, leading to the construction of an optimal model. To verify the impact of different training strategies on the model's Chinese mitten crab defect recognition and classification tasks, six different training strategies were designed using the improved SMP-EfficientVit model as the baseline model by modifying the loss function and optimiser, as shown in Table 3.

Table 3 Comparison of different training strategies

Strategy	Model	Optimiser	Loss Function	Batch size	Epochs	Learning rate
1	SMP-EfficientVit	Adam	Cross-entropy loss	16	100	0.001
2	SMP-EfficientVit	Adam	Focal loss	16	100	0.001
3	SMP-EfficientVit	Adam	Hinge loss	16	100	0.001
4	SMP-EfficientVit	SGD	Cross-entropy loss	16	100	0.001
5	SMP-EfficientVit	SGD	Focal loss	16	100	0.001
6	SMP-EfficientVit	SGD	Hinge loss	16	100	0.001

The evaluation results of the comparison test for different training strategies are shown in Table 4. Through experimental comparison, strategy one was selected, which uses the Adam optimiser (Kingma and Ba, 2015) combined with the cross-entropy loss function (Mao et al., 2023). This strategy demonstrated superior overall performance in the training process of the SMP-EfficientVit model. In terms of accuracy, precision, recall, and F1-score, strategy one achieved 97.34%, 97.35%, 97.34%, and 97.34%, respectively. These results demonstrate that the Adam optimiser has better convergence characteristics and stability compared to other optimisation algorithms in the parameter optimisation process of the SMP-EfficientVit model. Additionally, the cross-entropy loss function shows higher accuracy and efficiency in assessing the difference between model prediction outputs and actual labels compared to other loss functions. By minimising prediction errors, this loss function effectively encourages the model to produce clearer classification boundaries, enhancing the model's generalisation ability in the Chinese mitten crab defect detection task.

Table 4 Evaluation results of the comparison test for different training strategies

Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	97.34	97.35	97.34	97.34
2	95.27	95.32	95.29	95.30
3	94.64	94.65	94.59	94.62
4	95.42	95.57	95.39	95.48
5	93.74	93.81	93.73	93.77
6	92.69	92.70	92.68	92.69

3.2 Chinese mitten crab defect detection and classification results based on SMP-EfficientVit

A defect detection and classification model for Chinese mitten crabs based on the improved EfficientVit has been proposed. After experiments, the optimal training parameters and configurations of the model are shown in Table 5.

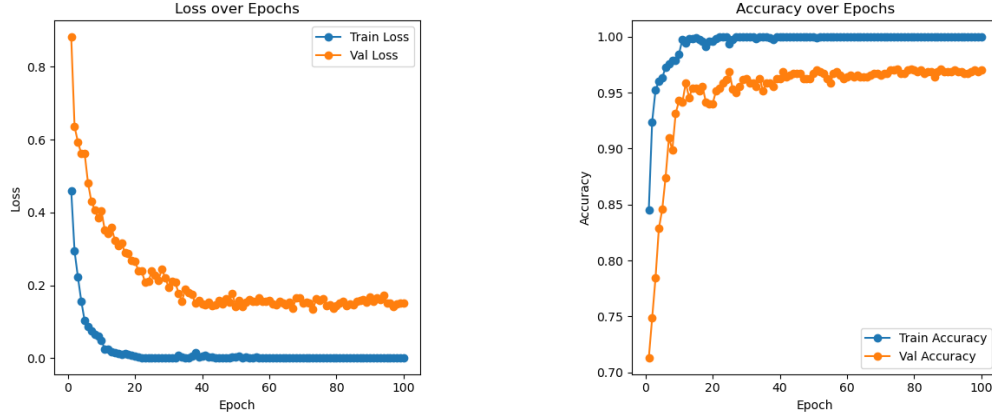
In Table 5, num_heads, router_dim, and top_k are key parameters of the MoH module, which together determine the efficiency and effectiveness of the model's attention mechanism. The num_heads parameter is set to 8, indicating that the model will use 8 different attention heads when processing features, which helps to capture the complex features of the input data from multiple perspectives. The router_dim parameter is set to 64, which determines how the router effectively selects and weights the outputs of different attention heads, enhancing the model's ability to capture key features. The top_k parameter is set to 1,

meaning that in each input token of the model, only the most relevant attention head is selected to participate in the final feature representation. This design aims to improve the model's computational efficiency while maintaining sensitivity to key information. The reduction and embed_dim are key parameters of the SE module. The embed_dim is 768, which represents the number of channels in the feature map processed by the SE module. A larger dimension helps to capture more detailed information. The reduction is 16, indicating the compression ratio of the feature map after global average pooling through two fully connected layers in the SE module. This parameter helps to reduce the number of parameters while maintaining the feature expression ability.

Table 5 Model training parameters and configurations

Parameter name	Parameter value	Parameter name	Parameter value
batch_size	16	step_size	10
epochs	100	gamma	0.5
num_classes	2	dropout_rate	0.4
num_heads	8	criterion	Cross-entropy loss
embed_dim	768	reduction	16
top_k	1	embed_dim	768
learning_rate	0.001	optimiser	Adam

After the parameter settings were completed, the training began. As the number of iterations increased, the model gradually stabilised. The loss curves and accuracy curves of the training set and validation set are shown in Figure 6. From Figure 6, it can be observed that as the training progressed, the accuracy of the model on the training set gradually increased and stabilised, while the loss value gradually decreased, indicating that the model learned effective feature representations during the training process. At the same time, the trend of accuracy and loss value on the validation set was similar to that on the training set, indicating that the model has good generalisation ability and can transfer the knowledge learned from the training data to unseen data. It is worth noting that the model's accuracy increased rapidly in the early stages of training, and then the rate of increase gradually slowed down. This may be due to the model quickly learning more obvious and common features in the initial stage, and as the training deepened, the model began to capture more subtle feature differences. The downward trend of the loss value also indicates that the model effectively reduced the prediction error during the optimisation process. In addition, the accuracy and loss value curves of the training set and validation set did not show any obvious divergent trends, which further verified the model's generalisation ability.

Figure 6 Loss and accuracy curves of the SMP-EfficientVit model on the training and validation sets (see online version for colours)

The trained SMP-EfficientVit model was applied to the testing set. The detection accuracy for defective crabs was 97.34%, with a precision of 98.28%, a recall of 97.65%, and an F1 score of 97.97%. For intact crabs, the detection accuracy was 97.34%, with a precision of 95.56%, a recall of 96.73%, and an F1 score of 96.15%. These metrics indicate a high level of performance, demonstrating that the SMP-EfficientVit model can effectively distinguish between defective and intact crabs. The results are summarised in Table 6.

Table 6 SMP-EfficientVit model test results

Items	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Defective crabs	97.34	98.28	97.65	97.97
Intact crabs	97.34	95.56	96.73	96.15
Overall data	97.34	97.35	97.34	97.34

The overall data in Table 6 refers to the model's aggregated performance metrics calculated by macro-averaging all test samples. This approach assigns equal weight to both 'defective crabs' and 'intact crabs' categories, ensuring a balanced evaluation of classification capability without bias toward the majority class. Accuracy remains consistent across all rows as it is class-agnostic, while precision, recall, and F1 scores reflect the macro-averaged results.

As shown in Table 6, the model performed exceptionally well in the detection of defective crabs, with precision and recall rates both approaching 98%, and an F1 score of 97.97%. This indicates high accuracy and reliability in predicting defective crabs. For the detection of intact crabs, although the precision was slightly lower than

that for defective crabs, the recall and F1 score remained at a high level, indicating good performance in the identification of intact crabs.

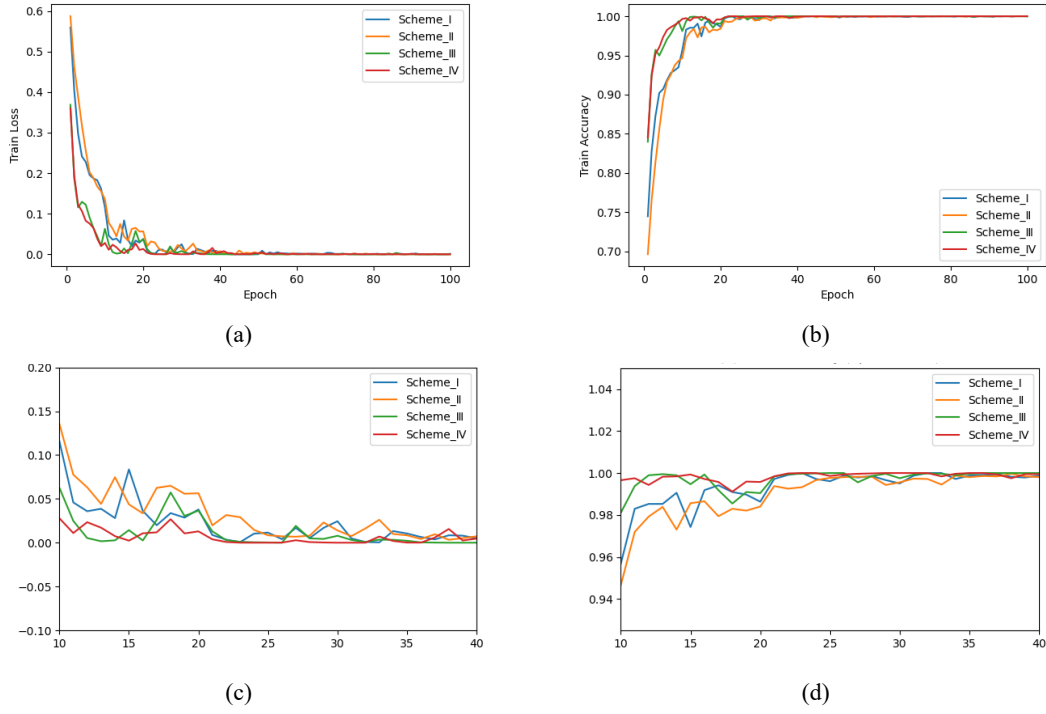
3.3 Ablation study

To thoroughly evaluate the effectiveness of the multiple improvements made to the EfficientVit model, an ablation study was conducted with detailed comparison analysis. The specific experimental schemes are shown in Table 7. Scheme I is the baseline model EfficientVit without any modifications. Scheme II introduces the SE module on this basis. Scheme III further incorporates the MoH module. Scheme IV replaces all ReLU activation functions with PReLU based on Scheme III.

Table 7 Ablation study schemes

Scheme	SE	MoH	PReLU	Batch size	Epochs	Learning rate
I				16	100	0.001
II	√			16	100	0.001
III	√	√		16	100	0.001
IV	√	√	√	16	100	0.001

Figures 7(a)–7(b) show the trends of accuracy and loss values for each scheme during the training process. As shown in the figure, as training progresses, the accuracy of each scheme gradually increases while the loss values gradually decrease, indicating that the model is continuously learning effective feature representations. Notably, Scheme IV exhibits more significant changes in accuracy and loss, suggesting that the introduced improvements effectively enhance the model's learning and generalisation abilities.

Figure 7 Training performance of each scheme in the ablation study, (a) training loss comparison (b) training accuracy comparison (c) training loss (epoch 10–40) (d) training accuracy (epoch 10–40) (see online version for colours)

Figures 7(c) and 7(d) are zoomed-in views of Figures 7(a) and 7(b), respectively, further displaying the changes in accuracy and loss values for Schemes I to IV during the early stages of training (from epoch 10 to 40). From these zoomed-in views, it is clear that Scheme IV has a faster accuracy improvement and a more rapid loss decrease compared to other schemes during the initial training stages. This further verifies the positive impact of the PReLU activation function in accelerating model convergence and enhancing model performance. Additionally, it indicates that the introduction of the SE and MoH modules enables the model to capture key features more quickly in the early stages of training, laying a solid foundation for subsequent training.

Table 8 presents the evaluation results of different schemes. As shown in Table 8, based on the baseline model EfficientVit, Scheme IV, which introduces the SE module, the MoH module, and replaces the activation function with PReLU, significantly improves the model's performance metrics. Specifically, compared to Scheme I, the accuracy, precision, recall, and F1 score of Scheme IV increased by 3.22%, 3.19%, 3.23%, and 3.29% respectively; compared to Scheme II, they increased by 2.10%, 2.10%, 2.10%, and 2.14% respectively; compared to Scheme III, they increased by 0.56%, 0.57%, 0.56%, and 0.56% respectively. These improvements demonstrate that Scheme IV has a significant advantage in enhancing model performance. The introduction of the SE and MoH modules strengthens the

model's ability to capture key features, while the PReLU activation function enhances the model's sensitivity to negative inputs and feature expression capabilities. Together, these factors contribute to the model's superior performance in the Chinese mitten crab defect detection task, highlighting the effectiveness of the proposed improvements.

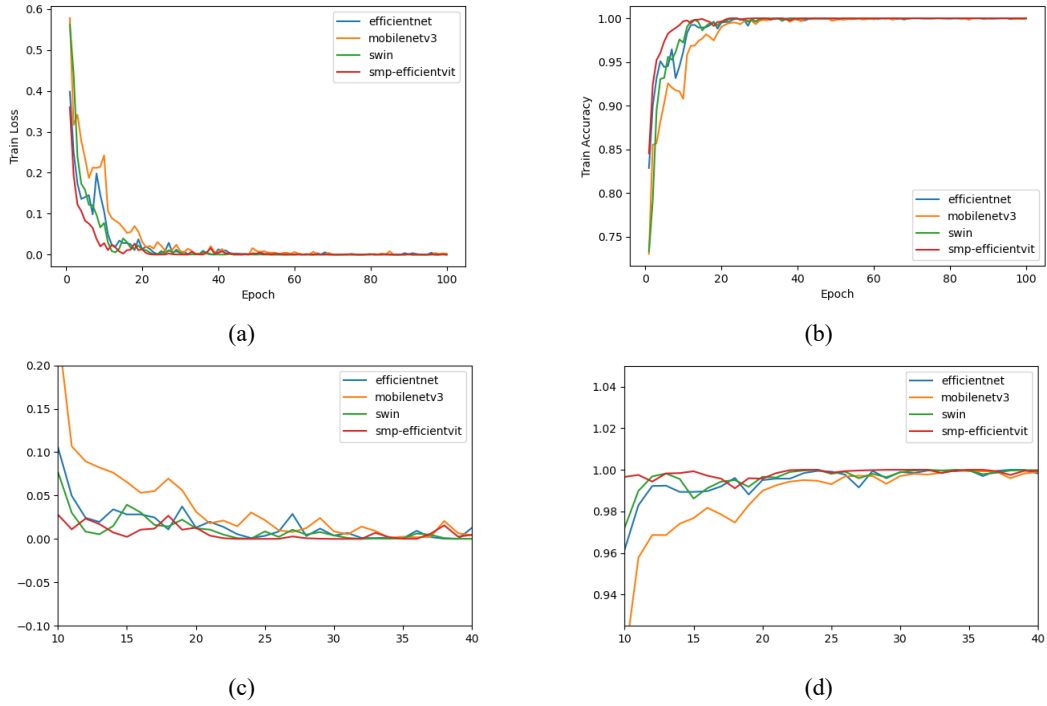
Table 8 Evaluation results of each scheme in the ablation study

<i>Scheme</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-score (%)</i>
I	94.12	94.16	94.12	94.05
II	95.24	95.25	95.24	95.20
III	96.78	96.78	96.78	96.78
IV	97.34	97.35	97.34	97.34

3.4 Comparison with different models

To fully verify the ability of the SMP-EfficientVit model for Chinese mitten crab defect detection and classification, it was compared with three other representative deep learning models: EfficientNet (Tan and Le, 2019), MobileNet V3 (Howard et al., 2019), and Swin transformer (Liu et al., 2021), using the same training strategy. The results are shown in Figure 8.

Figure 8 Training performance of each model, (a) training loss comparison (b) training accuracy comparison (c) training loss (epoch 10–40) (d) training accuracy (epoch 10–40) (see online version for colours)



Observing the overall trends in Figures 8(a)–8(b), the accuracy curve of the SMP-EfficientVit model during training is generally higher than those of the three traditional models being compared. This indicates that SMP-EfficientVit has a significant advantage in feature extraction and representation. The SMP-EfficientVit model also shows a faster trend in the decrease of loss values, reflecting its high efficiency in reducing prediction errors. In particular, in the zoomed-in views of the early training stages (epochs 10 to 40) in Figures 8(c)–8(d), the SMP-EfficientVit model exhibits a rapid increase in accuracy and a swift decrease in loss values. This further confirms the positive impact of the introduced SE module, MoH module, and PReLU activation function on enhancing the model's learning and generalisation abilities. These improvements enable the SMP-EfficientVit model to quickly capture key features in the early stages of training, laying a solid foundation for further training of the model.

The proposed SMP-EfficientVit model outperforms EfficientNet, MobileNet V3, and Swin Transformer in terms of accuracy, precision, recall, and F1 score, as shown in Table 9. Specifically, compared to EfficientNet, the SMP-EfficientVit model achieves improvements of 1.68%, 1.53%, 1.69%, and 1.61% in accuracy, precision, recall, and F1 score, respectively. Compared to MobileNet V3, the improvements are 2.24%, 2.24%, 2.02%, and 2.08%, respectively. Compared to Swin transformer, the improvements are 1.70%, 1.49%, 1.33%, and 1.37%, respectively. These results indicate that the SMP-EfficientVit model is effective in improving performance, mainly due to the enhanced ability to capture key features through the SE and MoH modules, and the increased sensitivity to negative inputs and feature

expression ability provided by the PReLU activation function.

Table 9 Comparison of SMP-EfficientVit with traditional network models

Model name	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EfficientNet	95.66	95.82	95.65	95.73
MobileNet V3	95.10	94.11	94.72	94.41
Swin transformer	95.64	95.85	96.01	95.93
SMP-EfficientVit	97.34	97.35	97.34	97.34

3.5 Limitations of the proposed method

Although the proposed SMP-EfficientVit model has demonstrated excellent performance in the Chinese mitten crab defect detection task, it still has some limitations. The image acquisition environment of the dataset constructed in this paper is relatively singular. Although data augmentation techniques have been used to simulate different lighting and background conditions, the complexity of the environment in practical applications may exceed the scope of experimental simulation. In the future, we will further train and test the model on more diverse datasets to improve its generalisation ability.

4 Conclusions

The SMP-EfficientVit model, developed for defect detection in Chinese mitten crabs, is based on an improved EfficientVit architecture. By constructing a dataset of intact

and defective crabs and incorporating advanced SE and MoH modules along with the PReLU activation function, the model achieves outstanding performance in defect classification tasks, with an accuracy of 97.34%, precision of 97.35%, recall of 97.34%, and F1 score of 97.34%. These results demonstrate the method's effectiveness and practical potential for automated quality control in aquaculture. The approach not only enhances detection efficiency but also significantly reduces several key human interference factors. For instance, it minimises subjective judgement variations caused by different inspectors' standards, reduces fatigue-induced errors during prolonged manual inspections, and avoids inconsistencies resulting from varying levels of professional skill. Additionally, the model's high accuracy decreases the likelihood of human error in defect identification, and its automated nature ensures consistent performance regardless of environmental conditions like lighting and background complexity. These improvements make it a scalable solution for boosting productivity and economic returns in the Chinese mitten crab industry.

Looking ahead, this method will be applied to real-world scenarios including automated sorting systems for rapid defective crab removal, early disease warning platforms to identify defects linked to environmental stressors, and standardised grading workflows that provide quantifiable metrics for commercial valuation. Future work will prioritise hardware-software co-design to deploy the lightweight model on edge devices and validate its robustness across diverse aquaculture environments with variable lighting conditions and complex backgrounds. Further integration with multi-modal sensing data, such as spectral imaging for shell integrity analysis, will expand its diagnostic capabilities, ultimately advancing intelligent aquaculture management systems.

Acknowledgements

This work was financially supported by Scientific Research Project of Liaoning Provincial Department of Education (JYTMS20230495, LJ212410158025), Special Fund for Basic Research Business Expenses of Undergraduate Universities in Liaoning Province (2024JBZ006), the Project of Liaoning Province Science and Technology Plan (2023JH1/10400043) and the Liaoning Social Science Fund Planning Project (L23BJY026).

Declarations

All authors declare that they have no conflicts of interest.

References

- Bai, S., Wang, S., Qin, D., Gao, L., Hao, Q., Du, N. and Wang, P. (2023) 'Elemental analysis of wild *Eriocheir sinensis*: determining the geographic origin and human health risk assessment', *Quality Assurance and Safety of Crops & Foods*, Vol. 15, No. 4, pp.133–142.
- Cao, S., Zhao, D., Liu, X. and Sun, Y. (2020) 'Real-time robust detector for underwater live crabs based on deep learning', *Computers and Electronics in Agriculture*, Vol. 172, No. 5, p.105339.
- Che, S., Li, S., Li, Q., Sun, Y., Zheng, Z., Nie, Z., Tang, Z., Wang, P., Gao, J. and Xu, G. (2022) 'Polyculture affects the growth, antioxidant status, nutrient content, and flavor of Chinese mitten crabs (*Eriocheir sinensis*) and largemouth bass (*Micropterus salmoides*)', *Fishes*, Vol. 7, No. 6, p.355.
- Chen, X., Zhang, Y., Li, D. and Duan, Q. (2023) 'Chinese mitten crab detection and gender classification method based on GMNet-YOLOv4', *Computers and Electronics in Agriculture*, Vol. 214, No. 11, p.108318.
- Cui, Y., Pan, T., Chen, S. and Zou, X. (2020) 'A gender classification method for Chinese mitten crab using deep convolutional neural network', *Multimedia Tools and Applications*, Vol. 79, No. 11, pp.7669–7684.
- Gu, H., Chen, M. and Gan, D. (2024) 'Gender identification of Chinese mitten crab juveniles based on improved faster R-CNN', *Applied Sciences*, Vol. 14, No. 2, p.908.
- He, K., Zhang, X., Ren, S. and Sun, J. (2015) 'Delving deep into rectifiers: surpassing human-level performance on ImageNet classification', *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December, pp.1026–1034.
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V. and Adam, H. (2019) 'Searching for MobileNetV3', *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October, IEEE, pp.1314–1324.
- Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E. (2018) 'Squeeze-and-excitation networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June, IEEE, pp.7132–7141.
- Jin, P., Zhu, B., Yuan, L. and Yan, S. (2024) *MoH: Multi-Head Attention as Mixture-of-Head Attention*, arXiv preprint arXiv:2410.11842.
- Kingma, D.P. and Ba, J. (2015) 'Adam: a method for stochastic optim ation', *3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA.
- Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H. and Yuan, Y. (2023) 'EfficientViT: memory efficient vision transformer with cascaded group attention', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June, IEEE, pp.12345–12355.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021) 'Swin transformer: hierarchical vision transformer using shifted windows', *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October, IEEE, pp.10012–10022.
- Luo, R., Jiang, T., Chen, X., Zheng, C., Liu, H. and Yang, J. (2019) 'Determination of geographic origin of Chinese mitten crab (*Eriocheir sinensis*) using integrated stable isotope and multi-element analyses', *Food Chemistry*, Vol. 274, No. 2, pp.1–7.
- Mao, A., Mohri, M. and Zhong, Y. (2023) 'Cross-entropy loss functions: theoretical analysis and applications', *Proceedings of the 40th International Conference on Machine Learning*, July, PMLR, Vol. 202, pp.23803–23828.
- Tan, M. and Le, Q. (2019) 'EfficientNet: rethinking model scaling for convolutional neural networks', *Proceedings of the 36th International Conference on Machine Learning (ICML)*, June, PMLR, Vol. 97, pp.6105–6114.

- Xue, J., Jiang, T., Chen, X., Liu, H. and Yang, J. (2022) 'Multi-mineral fingerprinting analysis of the Chinese mitten crab (*Eriocheir sinensis*) in Yangcheng Lake during the year-round culture period', *Food Chemistry*, Vol. 390, No. 11, p.133167.
- Zhang, J., Wang, S., Zhang, S., Li, J. and Sun, Y. (2023) 'Research on target detection and recognition algorithm of *Eriocheir sinensis* carapace', *Multimedia Tools and Applications*, Vol. 82, No. 23, pp.42527–42543.
- Zhang, R., Li, Y. and Guan, Z. (2024) 'Research on injection molded parts defect detection algorithm based on multiplicative feature fusion and improved attention mechanism', *Scientific Reports*, Vol. 14, No. 1, p.30864.
- Zhao, D., Liu, X., Sun, Y., Wu, R., Hong, J. and Ruan, C. (2019) 'Detection of underwater crabs based on machine vision', *Transactions of the Chinese Society of Agricultural Machinery*, Vol. 50, No. 3, pp.151–158.
- Zhu, P., Chen, C., Xu, B. and Lu, M. (2016) 'Research on freshness detection for Chinese mitten crab based on machine olfaction', in Tan, Y., Shi, Y. and Li, L. (Eds): *Advances in Swarm Intelligence: ICSI 2016, Lecture Notes in Computer Science*, Springer, Cham, Vol. 9713.