Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MSLG-Net: An Efficient Local–Global Feature Fusion Network for Shrimp Disease Recognition

Kaiyan Zhu
Dalian Ocean University
Yuhang Mao
15236610957@163.com

Dalian Ocean University
Shikai Zhang
Dalian Ocean University
Yuanbo Dai
Dalian Ocean University

Research Article

Keywords: shrimp disease recognition, lightweight neural network, local–global feature fusion, dual-label recognition, MSLG-Net

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11492966/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

MSLG-Net: An Efficient Local–Global Feature Fusion Network for Shrimp Disease Recognition

Kaiyan Zhu[1], Yuhang Mao[iD1*], Meini Wang[1], Shikai Zhang[1], Yuanbo Dai[1]

[1]School of Information Engineering, Dalian Ocean University, Dalian, Liaoning, China.

*Corresponding author(s). E-mail(s): 15236610957@163.com;

Abstract

Shrimp disease recognition remains challenging due to small lesion regions, subtle texture changes, similar disease appearances, complex backgrounds, and stringent lightweight deployment requirements. To address these issues, we propose MSLG-Net, a Multi-Scale Local–Global Network built upon EfficientNet-Lite0. Our framework integrates three synergistic modules: a Local Lesion Perception (LLP) module for enhancing subtle lesion signals, a Global Morphology Perception (GMP) module for capturing disease-related spatial morphology, and a Lesion-Guided Local–Global Gated Fusion (LGGF) module for adaptively fusing local cues with global semantic contexts under lesion-aware guidance. Experiments were conducted on two public single-label datasets, BD Fish & Shrimp Disease Dataset and TigerShrimpBD, and a self-constructed ShrimpMTD dual-label dataset. MSLG-Net consistently outperformed the EfficientNet-Lite0 baseline on the two public datasets. On ShrimpMTD, it achieved a Weighted Joint Accuracy of 0.8470, a Species Accuracy of 0.9343, and a Disease Accuracy of 0.8096. Ablation and visualization results further confirm the effectiveness of the proposed modules. Overall, MSLG-Net provides a feasible lightweight solution for shrimp disease recognition in resource-constrained aquaculture scenarios.

Keywords: shrimp disease recognition, lightweight neural network, local–global feature fusion, dual-label recognition, MSLG-Net

1 Introduction

Shrimp aquaculture is an important sector of global aquaculture, but intensive farming and complex aquatic environments make shrimp vulnerable to viral, bacterial, and environmentally induced diseases. White spot syndrome virus (WSSV) disease and yellow head disease are among the most destructive viral diseases in penaeid shrimp, often causing abnormal body coloration, changes in the cephalothorax, reduced vitality, and rapid mass mortality (Sánchez-Paz 2010; Munro and Owens 2007). Other common diseases, such as black gill disease and vibriosis, may also lead to gill discoloration, abnormal body color, and local tissue damage (Frischer et al. 2022; de Souza Valente and Wan 2021). Traditional diagnosis mainly relies on manual observation, empirical judgment, or laboratory testing. However, manual observation is subjective, while laboratory testing is usually time-consuming, costly, and difficult to apply for rapid on-site screening. Therefore, computer vision-based deep learning methods provide a promising approach for automatic and efficient shrimp disease recognition.

Deep learning has been widely applied in aquatic disease recognition, agricultural disease detection, and medical image-assisted diagnosis. Convolutional neural networks (CNNs) can automatically learn multi-level

visual features, including color, texture, edge, and morphological information, thereby reducing the dependence on manually designed features (Wu et al. 2025; Tamut et al. 2025). In shrimp disease recognition, CNN-based models, multi-scale convolutional networks, lightweight networks, transfer learning methods, and object detection models have been used for disease classification or lesion localization (Varma and Krishna 2022; Ramachandran et al. 2023; Aung et al. 2025; Fei et al. 2025). Meanwhile, public datasets such as ShrimpDiseaseImageBD, TigerShrimpBD, and the BD Fish & Shrimp Disease Dataset have provided useful data support for shrimp and aquatic disease image classification research (Islam and Rasel 2024; Ahmed and Farid 2025; Saon 2025).

While the aforementioned studies have substantially advanced the field, they fall short of addressing several practical challenges inherent to real-world shrimp disease recognition. First, disease-related regions are often small, weakly textured, and unclear in boundary, making key symptoms such as white spots, gill abnormalities, and local texture changes easy to overlook. Second, different diseases may present similar visual symptoms, such as body color changes, gill discoloration, and shell texture variations, which increases inter-class confusion. Third, real-world shrimp disease images usually involve limited samples, imbalanced categories, inconsistent shooting angles, and complex backgrounds. Finally, some high-accuracy models require large parameter counts and high computational costs, which limits their deployment in resource-constrained scenarios (Sharma et al. 2026; Jone et al. 2025). Therefore, it is necessary to improve the representation of local lesion details and global morphological semantics while maintaining low model complexity.

To address these problems, we propose MSLG-Net, a lightweight shrimp disease recognition model based on EfficientNet-Lite0 (Tan and Le 2019; Liu 2020). The proposed model introduces three complementary modules: a Local Lesion Perception (LLP) module for fine-grained lesion feature extraction, a Global Morphology Perception (GMP) module for global shrimp morphology modeling, and a Lesion-Guided Local–Global Gated Fusion (LGGF) module for adaptive local–global feature fusion. The main contributions of this study are summarized as follows:

- A lightweight EfficientNet-Lite0-based shrimp disease recognition model is constructed to improve disease-related feature representation while maintaining relatively low computational complexity.
- Three task-oriented modules, namely LLP, GMP, and LGGF, are designed to enhance local lesion perception, global morphology modeling, and adaptive local–global feature fusion for shrimp disease images.
- Experiments are conducted on two public single-label datasets and one self-constructed species–disease dual-label dataset to evaluate the proposed model, and the results demonstrate its effectiveness in both single-label disease classification and species–disease dual-label recognition.

2 Related Work

In recent years, deep learning has provided an effective technical route for shrimp disease image recognition. Early studies mainly focused on improving classification accuracy by enhancing feature extraction ability. For instance, SDNet combines clustering, feature selection, and deep convolutional classification to improve multi-class shrimp disease recognition (Varma and Krishna 2022), while DICNN introduces depthwise separable convolution, dense connections, and Inception-style multi-scale structures to strengthen the representation of white spot syndrome virus disease features (Ramachandran et al. 2023). These studies indicate that shrimp disease recognition benefits from multi-scale feature extraction and disease-region enhancement. However, most of them mainly emphasize classification performance, while the balance between fine-grained lesion sensitivity and model complexity remains insufficiently considered.

To further improve disease-related feature representation, attention mechanisms and feature fusion strategies have been introduced into shrimp disease recognition and related visual classification tasks. Models such as E-DarkNet53 and ConvNeXt-CBAM, where CBAM denotes the Convolutional Block Attention Module, show that multi-branch feature fusion and attention-guided enhancement can improve the discrimination of shrimp disease or health conditions (Lou et al. 2026; Le et al. 2025). Meanwhile, general attention modules such as Squeeze-and-Excitation (SE), CBAM, and Efficient Channel Attention (ECA) provide useful mechanisms for strengthening channel or spatial responses (Woo et al. 2018; Hu et al. 2018; Wang et al. 2020). These methods help models focus on more informative regions, but most existing designs still rely on either local attention enhancement or global feature aggregation separately. The adaptive collaboration between local lesion cues and global morphological semantics has not been sufficiently explored, especially under lightweight constraints.

Lightweight neural networks have become increasingly important for practical aquaculture applications, where recognition models are expected to run on devices with limited computational resources. MobileNetV3, MobileNetV4, and GhostNet reduce computational cost through architecture optimization or cheap feature generation operations (Howard et al. 2019; Qin et al. 2025; Han et al. 2020). FeatherNetX further demonstrates that task-specific lightweight design can achieve a favorable balance between shrimp disease classification accuracy and model efficiency (Sharma et al. 2026). Nevertheless, lightweight models may weaken the representation of subtle lesion patterns during compression or downsampling, which is unfavorable for diseases with small spots, weak texture changes, or similar visual symptoms. In addition, most existing shrimp disease recognition methods are designed for single-label classification, while species–disease dual-label recognition remains less investigated. Therefore, we build MSLG-Net upon EfficientNet-Lite0, aiming to enhance local lesion perception, global morphology modeling, and adaptive local–global fusion within a compact lightweight framework.

3 MSLG-Net Model

3.1 Overall Architecture

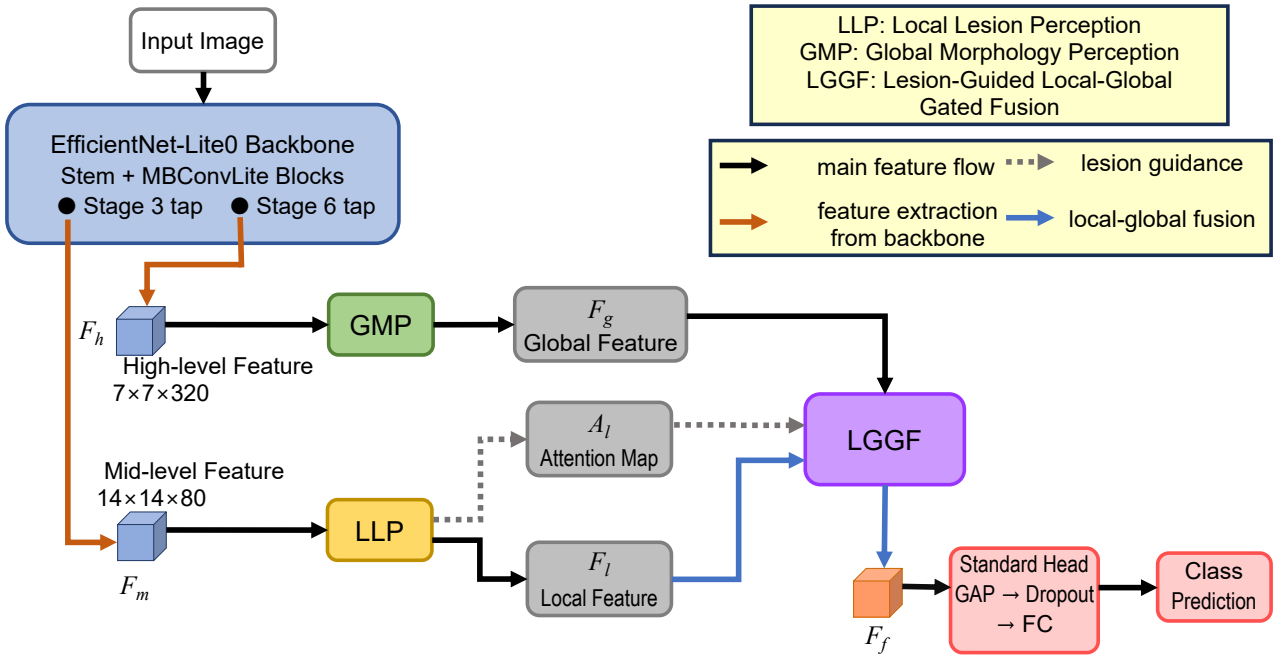


Fig. 1 Overall architecture of MSLG-Net. Solid arrows indicate feature flow; the dashed arrow denotes lesion guidance. GAP denotes global average pooling and FC denotes a fully connected layer

As shown in Fig. 1, MSLG-Net is constructed on EfficientNet-Lite0 and integrates three task-oriented components: Local Lesion Perception (LLP), Global Morphology Perception (GMP), and Lesion-Guided Local–Global Gated Fusion (LGGF). The backbone is responsible for extracting hierarchical visual features, while these components further strengthen fine-grained lesion representation, global morphological modeling, and adaptive local–global feature integration.

Specifically, the mid-level feature from the backbone is fed into the LLP module to generate local lesion features and a local attention map, whereas the high-level feature is fed into the GMP module to obtain global morphological features. The LGGF module then uses the local attention map to recalibrate the global feature and adaptively fuse local and global information through a gated mechanism. Finally, the fused feature is passed to the classification head for prediction. Let the input image be denoted as $\mathbf{X}$:

$$\mathbf{X} \in \mathbb{R}^{B\times 3\times H\times W} \tag{1}$$

Here, B represents the batch size, 3 denotes the red, green, and blue (RGB) channels, and H and W denote the height and width of the input image, respectively.

The input image is first processed by the EfficientNet-Lite0 backbone to extract hierarchical features, which can be represented as:

$$\{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_i, \cdots, \mathbf{F}_n\} = \text{Backbone}(\mathbf{X}) \tag{2}$$

Here, $\mathbf{F}_i$ denotes the feature map output from the i-th stage of the backbone network.

A mid-level feature $\mathbf{F}_m$ and a high-level feature $\mathbf{F}_h$ are selected from the backbone as the inputs of the subsequent modules:

$$\mathbf{F}_m = \mathbf{F}_{s_m}, \qquad \mathbf{F}_h = \mathbf{F}_{s_h} \tag{3}$$

Here, s_m denotes the stage of the selected mid-level feature, and s_h denotes the stage of the selected high-level feature. The mid-level feature $\mathbf{F}_m$ retains more spatial details and is mainly used to describe body surface white spots, gill color changes, and local texture abnormalities. The high-level feature $\mathbf{F}_h$ contains stronger semantic information and is mainly used to describe global shrimp morphology, body color distribution, and contextual relationships related to disease regions.

The mid-level feature $\mathbf{F}_m$ is fed into the LLP module to obtain the local lesion feature $\mathbf{F}_l$ and the local attention map $\mathbf{A}_l$:

$$\mathbf{F}_l, \mathbf{A}_l = \text{LLP}\left(\mathbf{F}_m\right) \tag{4}$$

Meanwhile, the high-level feature ($\mathbf{F}_h$) is fed into the GMP module to obtain the global morphological feature $\mathbf{F}_g$:

$$\mathbf{F}_g = \text{GMP}\left(\mathbf{F}_h\right) \tag{5}$$

After obtaining the local lesion feature and the global morphological feature, lesion-guided recalibration and lightweight gated fusion are integrated into the LGGF module. The module takes the local lesion feature $\mathbf{F}_l$, the local lesion attention map $\mathbf{A}_l$, and the global morphological feature $\mathbf{F}_g$ as inputs. It first uses the local attention map to guide the global feature to focus on disease-related regions, and then adaptively fuses the local and global features through a gating mechanism to obtain the fused feature $\mathbf{F}_f$:

$$\mathbf{F}_f = \text{LGGF}\left(\mathbf{F}_l, \mathbf{A}_l, \mathbf{F}_g\right) \tag{6}$$

Here, $\mathbf{F}_f$ denotes the fused local–global feature. Finally, the fused $\mathbf{F}_f$ feature is fed into the standard classification head to obtain the final prediction:

$$\widehat{y} = \text{Classifier}\left(\mathbf{F}_f\right) \tag{7}$$

Here, the classifier consists of global average pooling (GAP), dropout, and a fully connected (FC) layer, and $\widehat{y}$ denotes the final prediction of the model. For ordinary single-label shrimp disease datasets, $\widehat{y}$ represents the predicted disease category. For the ShrimpMTD dual-label task, MSLG-Net can be used as a shared feature extraction network and further connected to a multi-task classification head to output the shrimp species, disease type, and joint species–disease category.

Therefore, the overall forward propagation process of MSLG-Net can be summarized as follows:

$$\mathbf{X} \rightarrow \text{Backbone}(\mathbf{X}) \rightarrow (\mathbf{F}_m, \mathbf{F}_h) \tag{8}$$

$$\mathbf{F}_m \rightarrow \text{LLP} \rightarrow (\mathbf{F}_l, \mathbf{A}_l) \tag{9}$$

$$\mathbf{F}_h \rightarrow \text{GMP} \rightarrow \mathbf{F}_g \tag{10}$$

$$(\mathbf{F}_l, \mathbf{A}_l, \mathbf{F}_g) \rightarrow \text{LGGF} \rightarrow \mathbf{F}_f \tag{11}$$

$$\mathbf{F}_f \rightarrow \text{Classifier} \rightarrow \widehat{y} \tag{12}$$

The overall architecture enables MSLG-Net to jointly exploit mid-level lesion details and high-level morphological semantics for shrimp disease recognition.

3.2 EfficientNet-Lite0 Backbone

The backbone network plays a fundamental role in extracting discriminative visual features while controlling the overall computational cost of the model. The EfficientNet family uses compound scaling to jointly adjust network depth, width, and input resolution, achieving strong performance under relatively low computational cost (Tan and Le 2019). EfficientNet-Lite further simplifies the original EfficientNet design by removing the Squeeze-and-Excitation module and replacing the Swish activation function with a rectified linear unit capped at 6 (ReLU6), making it more suitable for lightweight deployment scenarios (Liu 2020).

In the proposed model, the EfficientNet-Lite0 backbone consists of a stem layer and multiple MBConvLite blocks. The stem layer uses a 3×3 convolution for initial feature extraction, while the MBConvLite blocks

progressively perform spatial downsampling and channel transformation. Each MBConvLite block follows the inverted residual and depthwise separable convolution design (Sandler et al. 2018), mainly including 1×1 channel expansion, depthwise convolution, 1×1 channel projection, and an optional residual connection. For an input feature $\mathbf{F}_{\text{in}}$, the computation process can be expressed as:

$$\mathbf{F}_e = \delta\left(\text{BN}\left(\text{Conv}_{1\times1}\left(\mathbf{F}_{\text{in}}\right)\right)\right) \tag{13}$$

$$\mathbf{F}_d = \delta\left(\text{BN}\left(\text{DWConv}_{k\times k}\left(\mathbf{F}_e\right)\right)\right) \tag{14}$$

$$\mathbf{F}_p = \text{BN}\left(\text{Conv}_{1\times1}\left(\mathbf{F}_d\right)\right) \tag{15}$$

$$\mathbf{F}_{\text{out}} = \mathbf{F}_p + \mathbf{F}_{\text{in}} \tag{16}$$

Here, $\text{Conv}_{1\times1}$ denotes pointwise convolution, $\text{DWConv}_{k\times k}$ denotes depthwise convolution with a $k\times k$ kernel, BN denotes batch normalization, and δ denotes the ReLU6 activation function.

When the input and output channels are different or the stride is not equal to 1, the residual connection is not enabled, and the output is computed as:

$$\mathbf{F}_{\text{out}} = \mathbf{F}_p \tag{17}$$

Different from the original EfficientNet-Lite0 classifier that mainly relies on the final-layer feature, MSLG-Net uses its mid-level and high-level features for local lesion perception and global morphology modeling, respectively, thereby providing multi-level representations for subsequent local–global feature fusion.

3.3 Local Lesion Perception Module

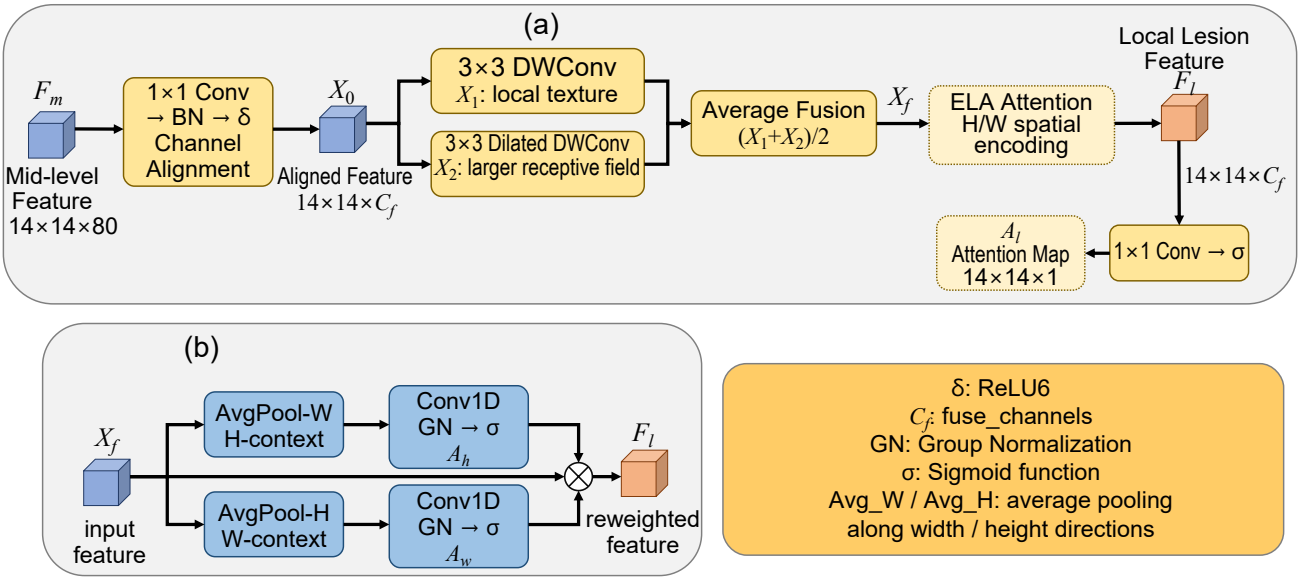


Fig. 2 Structure of the LLP module: (a) local lesion perception pipeline; (b) Efficient Location Attention (ELA)

Key lesion patterns in shrimp disease images, such as body surface white spots, abnormal gill coloration, local red spots, and texture changes, are often small and easily weakened during backbone downsampling. To preserve such fine-grained disease cues, we introduce the Local Lesion Perception (LLP) module after the mid-level feature stage of EfficientNet-Lite0, as shown in Fig. 2. Let the mid-level feature output from Stage 3 of the backbone be denoted as $\mathbf{F}_m$, whose size is:

$$\mathbf{F}_m \in \mathbb{R}^{14\times14\times80} \tag{18}$$

Here, 14×14 denotes the spatial size of the feature map, and 80 denotes the number of channels.

The LLP module applies a 1×1 convolution to align the channel dimension of the mid-level feature and map it to the fusion channel number C_f:

$$\begin{aligned}&\mathbf{X}_0 = \delta\left(\text{BN}\left(\text{Conv}_{1\times1}\left(\mathbf{F}_m\right)\right)\right),\\&\mathbf{X}_0 \in \mathbb{R}^{14\times14\times C_f}\end{aligned} \tag{19}$$

This operation unifies the feature channel dimension and reduces the computational complexity of subsequent local feature extraction. Then, two lightweight

branches are used to extract local lesion features. The standard 3×3 depthwise convolution branch captures local textures, edges, and small-scale lesion patterns:

$$\mathbf{X}_1 = \mathrm{DWConv}_{3\times3}(\mathbf{X}_0) \tag{20}$$

The LLP module introduces a dilated 3×3 depthwise convolution branch to enlarge the local receptive field without significantly increasing the number of parameters. This enables the model to perceive lesion regions together with their surrounding contextual information. Dilated convolution can effectively enlarge the receptive field and has been widely used in multi-scale context modeling tasks (Yu and Koltun 2015):

$$\mathbf{X}_2 = \mathrm{DWConv}^{d}_{3\times3}(\mathbf{X}_0) \tag{21}$$

Here, $\mathrm{DWConv}^{d}_{3\times3}$ denotes depthwise convolution with dilation rate. The outputs of the two branches are averaged to obtain the fused local feature:

$$\mathbf{X}_f = \frac{\mathbf{X}_1 + \mathbf{X}_2}{2} \tag{22}$$

Here, $\mathbf{X}_f$ denotes the fused local feature, which contains both fine-grained texture information and local contextual cues.

After obtaining the fused local feature, the LLP module introduces the Efficient Location Attention (ELA) mechanism to further enhance lesion-region responses. ELA models local positional information through one-dimensional spatial encoding along the height and width directions, enabling the feature map to emphasize key regions with low computational complexity (Xu et al. 2025). Average pooling is first performed along the width direction to obtain the height-wise contextual feature:

$$\mathbf{H}_{\mathrm{ctx}} = \mathrm{Avg}_W(\mathbf{X}_f) \tag{23}$$

Meanwhile, average pooling is performed along the height direction to obtain the width-wise contextual feature:

$$\mathbf{W}_{\mathrm{ctx}} = \mathrm{Avg}_H(\mathbf{X}_f) \tag{24}$$

Here, Avg_W denotes average pooling along the width direction, and Avg_H denotes average pooling along the height direction.

The two contextual features are then processed by 1D convolution, group normalization, and sigmoid activation to generate height-wise and width-wise attention weights:

$$\mathbf{A}_h = \sigma(\mathrm{GN}(\mathrm{Conv1D}(\mathbf{H}_{\mathrm{ctx}}))) \tag{25}$$

$$\mathbf{A}_w = \sigma(\mathrm{GN}(\mathrm{Conv1D}(\mathbf{W}_{\mathrm{ctx}}))) \tag{26}$$

Here, Conv1D denotes one-dimensional convolution, GN denotes group normalization, and σ denotes the sigmoid activation function.

Finally, the two directional attention weights are used to reweight the fused local feature $\mathbf{X}_f$, producing the local lesion feature $\mathbf{F}_l$:

$$\mathbf{F}_l = \mathbf{X}_f \odot \mathbf{A}_h \odot \mathbf{A}_w \qquad \mathbf{F}_l \in \mathbb{R}^{14\times14\times C_f} \tag{27}$$

Here, $\odot$ denotes element-wise multiplication, and $\mathbf{F}_l$ denotes the enhanced local lesion feature.

To provide lesion-guided information for the subsequent LGGF module, a local lesion attention map is generated through a 1×1 convolution and a sigmoid function:

$$\mathbf{A}_l = \sigma(\mathrm{Conv}_{1\times1}(\mathbf{F}_l)) \qquad \mathbf{A}_l \in \mathbb{R}^{14\times14\times1} \tag{28}$$

Here, $\mathbf{A}_l$ denotes the local lesion attention map, whose values range from 0 to 1. It reflects the response intensity of local lesion regions and is used to guide the recalibration of high-level global features in the LGGF module. Therefore, the overall output of the LLP module can be expressed as:

$$\mathbf{F}_l, \mathbf{A}_l = \mathrm{LLP}(\mathbf{F}_m) \tag{29}$$

Through this module, MSLG-Net supplements the lightweight backbone with stronger local detail representation, thereby improving its ability to recognize fine-grained shrimp disease features.

3.4 Global Morphology Perception Module

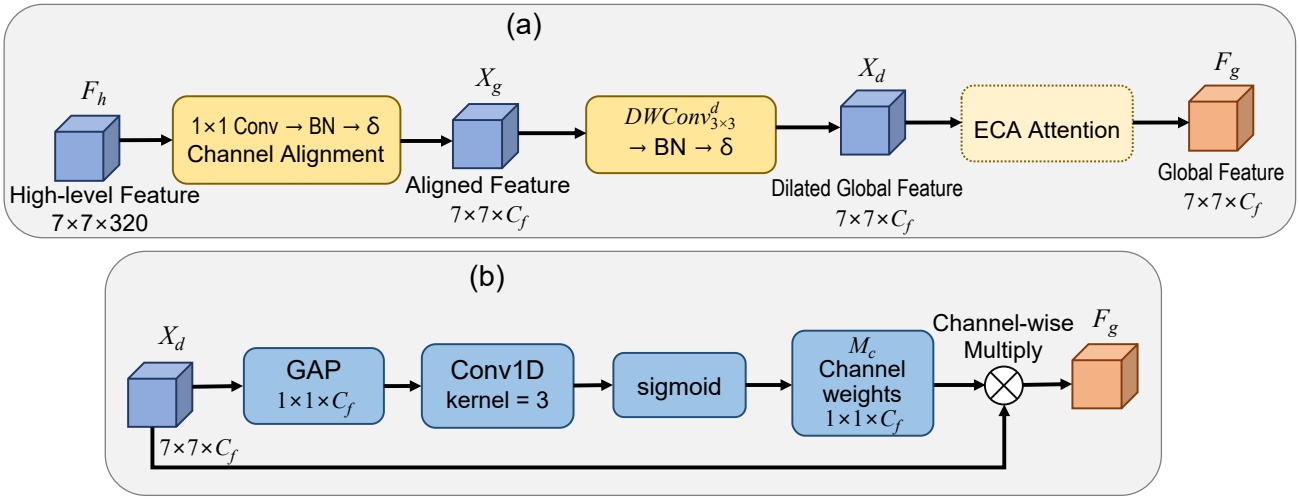


Fig. 3 Structure of the GMP module: (a) global morphology perception pipeline; (b) Efficient Channel Attention (ECA)

Shrimp disease recognition depends not only on local lesion details but also on global morphological cues, such as body structure, color distribution, and large-scale disease-related appearance. For example, black gill disease often causes darkened gill regions and cephalothorax discoloration, while vibriosis may be accompanied by abnormal body coloration and local red spots. To capture such global disease-related information, we introduce the Global Morphology Perception (GMP) module after the high-level semantic feature stage, as shown in Fig. 3.

Let the high-level feature output from Stage 6 of the EfficientNet-Lite0 backbone be denoted as $\mathbf{F}_h$, whose size is:

$$\mathbf{F}_h \in \mathbb{R}^{7\times7\times320} \tag{30}$$

Here, 7×7 denotes the spatial size of the high-level feature map, and 320 denotes the number of channels.

The GMP module first applies a 1×1 convolution to compress and align the channel dimension of $\mathbf{F}_h$, mapping it to the fusion channel number C_f:

$$\mathbf{X}_g = \delta\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}\left(\mathbf{F}_h\right)\right)\right) \qquad \mathbf{X}_g \in \mathbb{R}^{7\times7\times C_f} \tag{31}$$

This operation reduces the channel dimension of the high-level feature and aligns it with the local feature produced by the LLP module. Then, a dilated 3×3 depthwise convolution followed by batch normalization and ReLU6 activation is used to extract global morphological features:

$$\mathbf{X}_d = \delta\left(\mathrm{BN}\left(\mathrm{DWConv}^d_{3\times3}\left(\mathbf{X}_g\right)\right)\right) \tag{32}$$

Here, $\mathbf{X}_d$ denotes the global morphological feature after dilated depthwise convolution, batch normalization, and ReLU6 activation.

To further enhance disease-related channel responses, the GMP module introduces the Efficient Channel Attention (ECA) mechanism after the dilated depthwise convolution. ECA models local cross-channel interactions through global average pooling and one-dimensional convolution, avoiding the dimensionality reduction operation used in traditional channel attention mechanisms. Therefore, ECA can improve feature representation with only a small number of additional parameters (Wang et al. 2020). Global average pooling is applied to the feature map $\mathbf{X}_d$ to obtain the channel descriptor $\mathbf{z}$:

$$\mathbf{z} = \mathrm{GAP}\left(\mathbf{X}_d\right) \tag{33}$$

The channel descriptor is then processed by one-dimensional convolution and sigmoid activation to generate channel attention weights:

$$\mathbf{M}_c = \sigma\left(\mathrm{Conv1D}(\mathbf{z})\right) \tag{34}$$

Here, $\mathbf{M}_c$ denotes the channel attention weights.

Finally, the channel attention weights are broadcast to the same spatial size as $\mathbf{X}_d$ and multiplied with

$\mathbf{X}_d$ channel by channel to obtain the refined global morphological feature:

$$\mathbf{F}_g = \mathbf{X}_d \odot \mathbf{M}_c \qquad \mathbf{F}_g \in \mathbb{R}^{7\times7\times C_f} \tag{35}$$

The resulting feature preserves high-level semantic information and strengthens channel responses related to shrimp morphology, body color changes, and disease categories. Therefore, the overall output of the GMP module can be expressed as:

$$\mathbf{F}_g = \mathrm{GMP}\left(\mathbf{F}_h\right) \tag{36}$$

Thus, GMP provides compact global morphological representations for subsequent local–global fusion.

3.5 Lesion-Guided Local–Global Gated Fusion Module

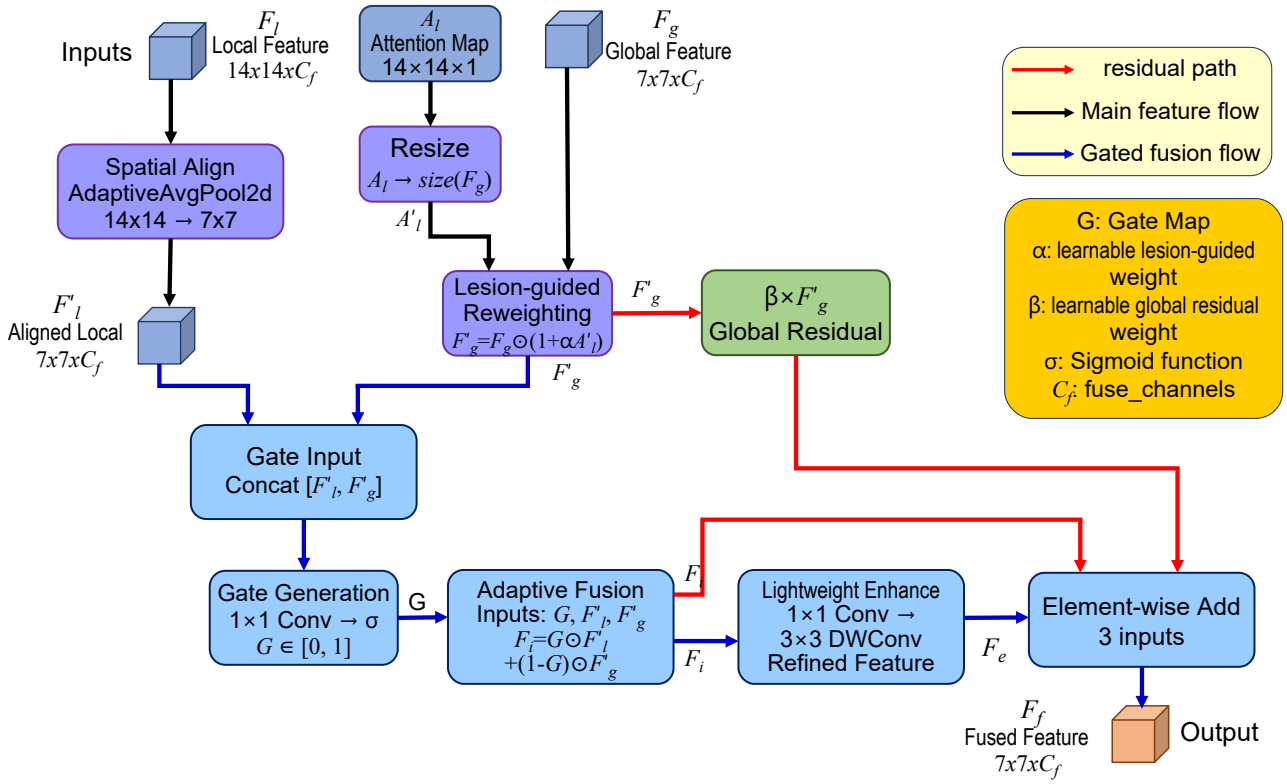


Fig. 4 Structure of the LGGF module. Black solid arrows indicate the main feature flow, red solid arrows indicate residual connections, and blue solid arrows indicate gated fusion

Although the LLP and GMP modules provide local lesion features and global morphological features, direct fusion by simple addition or channel concatenation may introduce cross-scale interference and increase computational cost. Existing attention and feature fusion studies have shown that channel, spatial, or region-level guidance can strengthen the representation of discriminative regions (Woo et al. 2018; Hu et al. 2018). Based on this idea, we design the Lesion-Guided Local–Global Gated Fusion (LGGF) module to integrate lesion-guided recalibration and lightweight gated fusion, as shown in Fig. 4.

Since the spatial sizes of the local attention map and the global feature may be different, the local attention map is first resized to match the spatial scale of the global feature:

$$\mathbf{A}'_l = \mathrm{Resize}\left(\mathbf{A}_l, \mathrm{size}\left(\mathbf{F}_g\right)\right) \tag{37}$$

Here, $\mathbf{A}'_l$ denotes the resized local lesion attention map. Then, $\mathbf{A}'_l$ is used to perform lesion-guided recalibration on the global morphological feature:

$$\mathbf{F}'_g = \mathbf{F}_g \odot \left(1 + \alpha \mathbf{A}'_l\right) \tag{38}$$

Here, α is a learnable weight coefficient used to control the enhancement strength of the local lesion attention on the global feature. The residual-style recalibration preserves the original global feature while

emphasizing disease-related regions, preventing useful contextual information from being overly suppressed.

After global feature recalibration, the local lesion feature is resized to match the spatial size of the recalibrated global feature:

$$\mathbf{F}_l' = \text{ Resize}\left(\mathbf{F}_l, \text{size}\left(\mathbf{F}_g'\right)\right) \tag{39}$$

The resized local lesion feature and the recalibrated global feature are then concatenated and passed through a 1×1 convolution followed by sigmoid activation to generate the gate weight ($\mathbf{G}$):

$$\mathbf{G} = \sigma\left(\text{Conv}_{1\times1}\left(\left[\mathbf{F}_l', \mathbf{F}_g'\right]\right)\right) \qquad \mathbf{G} \in [0,1] \tag{40}$$

Here, $[\cdot,\cdot]$ denotes concatenation along the channel dimension.

The local and global features are fused through gated weighting:

$$\mathbf{F}_i = \mathbf{G} \odot \mathbf{F}_l' + (1-\mathbf{G}) \odot \mathbf{F}_g' \tag{41}$$

Here, $\mathbf{F}_i$ denotes the initially fused local–global feature. A larger value of $\mathbf{G}$ indicates stronger reliance on local lesion details, while a smaller value of $\mathbf{G}$ gives more weight to global morphological semantics. In this way, LGGF adaptively adjusts the contribution of local and global information according to the disease appearance of the input image.

To further enhance local interaction while maintaining lightweight complexity, LGGF applies a 1×1 convolution and a 3×3 depthwise convolution after gated fusion:

$$\mathbf{F}_e = \text{DWConv}_{3\times3}\left(\text{Conv}_{1\times1}\left(\mathbf{F}_i\right)\right) \tag{42}$$

Here, the 1×1 convolution is used for channel transformation, and the 3×3 depthwise convolution is used to model local spatial relationships with low computational cost.

Finally, the enhanced feature, the initially fused feature, and the recalibrated global feature are combined through residual fusion to obtain the final output:

$$\mathbf{F}_f = \ \mathbf{F}_e + \mathbf{F}_i + \beta\mathbf{F}_g' \tag{43}$$

Here, $\mathbf{F}_f$ denotes the final fused feature, and β is a learnable global semantic residual coefficient. This residual term preserves high-level global semantic information and prevents the local lesion branch from weakening global morphology-related representations.

Therefore, the overall process of the LGGF module can be expressed as:

$$\mathbf{F}_f = \text{LGGF}\left(\mathbf{F}_l, \mathbf{A}_l, \mathbf{F}_g\right) \tag{44}$$

In summary, the LGGF module adaptively integrates local lesion features and global morphological features through lesion-guided recalibration, gated fusion, and lightweight residual enhancement.

3.6 Classification Output and Loss Function

After local–global fusion, the fused feature $\mathbf{F}_f$ is used for final prediction. For single-label disease classification, a standard classification head consisting of global average pooling, dropout, and a fully connected layer is adopted:

$$\mathbf{o} = \text{FC}\left(\text{Dropout}\left(\text{GAP}\left(\mathbf{F}_f\right)\right)\right) \tag{45}$$

Here, GAP denotes global average pooling, FC denotes the fully connected layer, and $\mathbf{o}$ denotes the output logits of the model. The predicted disease category is obtained from $\mathbf{o}$.

For single-label classification, the model is optimized using cross-entropy loss:

$$L_{\text{cls}} = -\sum_{i=1}^{N} y_i \log\left(\widehat{y}_i\right) \tag{46}$$

Here, N denotes the number of classes, y_i denotes the one-hot encoded ground-truth label, and $\widehat{y}_i$ denotes the predicted probability of class i.

For the ShrimpMTD dual-label task, MSLG-Net serves as a shared feature extractor. The fused feature is first processed by global average pooling to obtain a feature vector $\mathbf{v}$. Based on this feature vector, a species prediction head, a disease prediction head, and a joint category prediction head are constructed:

$$\mathbf{v} = \text{GAP}\left(\mathbf{F}_f\right) \tag{47}$$

$$\mathbf{o}_s, \mathbf{o}_d, \mathbf{o}_j = \text{Head}_s(\mathbf{v}),\ \text{Head}_d(\mathbf{v}), \text{Head}_j(\mathbf{v}) \tag{48}$$

Here, $\mathbf{o}_s$ denotes the shrimp species prediction logits, $\mathbf{o}_d$ denotes the disease type prediction logits, and $\mathbf{o}_j$ denotes the joint category prediction logits.

Let the number of disease categories be N_d, the ground-truth species label of a sample be y_s, and the ground-truth disease label be y_d. The joint category label can be expressed as:

$$y_j = \ y_s \times N_d + y_d \tag{49}$$

During training, the species loss L_s, disease loss L_d, and joint category loss L_j are calculated separately. Since disease recognition is the primary objective of this study, a higher weight is assigned to the disease classification loss. The total loss for the dual-label task is defined as:

$$L_{\text{total}} = 0.20L_s + 0.70L_d + 0.10L_j \tag{50}$$

Here, L_s, L_d, and L_j are all computed using cross-entropy loss. This weighted multi-task loss enables the model to jointly optimize species recognition, disease recognition, and joint category prediction, while placing more emphasis on disease-type discrimination.

In summary, the proposed classification and loss design supports both single-label disease classification and species–disease dual-label recognition, enabling MSLG-Net to learn disease-oriented discriminative features from shared local–global representations.

4 Experimental Design

4.1 Experimental Environment and Parameter Settings

Table 1 Training parameter settings

Parameter	**Setting**
Input image size	224×224
Batch Size	16
Training epochs	100 epochs for public datasets; 150 epochs for the self-constructed dual-label dataset
Initial learning rate	2×10^{-4}
Optimizer	AdamW
Weight decay	0.0001

All experiments were conducted on the Windows 11 operating system. Model training and testing were implemented using Python 3.9 and the PyTorch 2.8.0 deep learning framework. In terms of hardware configuration, the central processing unit (CPU) was an AMD Ryzen 7 5800X, and the graphics processing unit (GPU) was an NVIDIA GeForce RTX 3080 Ti. CUDA 12.6 was used for GPU acceleration. The main training settings are listed in Table 1.

To ensure fair comparison, all models were trained and evaluated under the same input size, data split, optimizer, and main hyperparameter settings. In addition, each model was trained and evaluated three times using different random seeds to reduce the influence of random initialization, data shuffling, and training stochasticity. The final results are reported as the mean ± standard deviation over the three runs.

4.2 Datasets and Preprocessing

Three datasets were used to evaluate MSLG-Net: the BD Fish & Shrimp Disease Dataset, TigerShrimpBD, and the ShrimpMTD dataset. The first two datasets were used for single-label disease classification, while ShrimpMTD was used for species–disease dual-label recognition.

For the BD Fish & Shrimp Disease Dataset, all 11 categories were used, and the original labels were retained. For TigerShrimpBD, the split was performed at the original-image level to avoid data leakage caused by augmented versions of the same original image appearing in different subsets.

ShrimpMTD is a self-constructed species–disease dual-label dataset consisting of three shrimp species, namely *Litopenaeus vannamei*, *Macrobrachium rosenbergii*, and *Penaeus monodon*, and four health or disease states, including healthy, white spot syndrome, black gill disease, and vibriosis. These labels form 12 species–disease combination categories, with a total of 2,090 images.

The ShrimpMTD images were collected from field-captured samples, publicly accessible online sources, and training-set augmentation. Field images were captured in aquaculture or sampling scenarios, while online images were collected from publicly accessible aquaculture, disease popularization, or image information webpages. During data cleaning, images with low clarity, incomplete subjects, high duplication, or indistinguishable disease features were removed. Training-set augmented images were generated only from the training split using conventional operations such as random flipping, rotation, and color jittering, and were not included in the validation or test sets. The remaining images were manually annotated according to shrimp morphology, body color variation, gill-region appearance, and body surface lesion patterns, resulting in species labels, disease labels, and joint species–disease labels. The class distribution of ShrimpMTD is shown in Fig. 5.

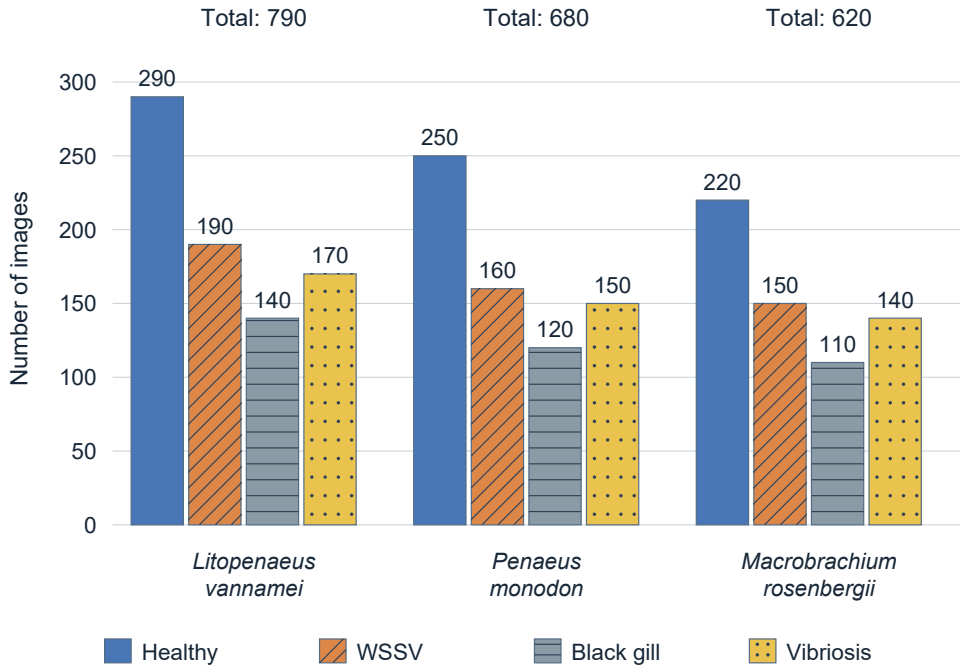


Fig. 5 Class distribution of the ShrimpMTD dataset. Bar labels indicate image counts, and the totals above the groups give the numbers of images for each species

All images were resized to 224 × 224 before being fed into the model. For the public single-label datasets, images were read by category and split into training, validation, and test sets using stratified sampling with a ratio of 7:2:1. Random horizontal flipping, random rotation, and color jittering were applied only during training, while validation and testing used only resizing and normalization. For ShrimpMTD, images were organized in a "shrimp species/disease type/image" hierarchy, allowing species, disease, and joint labels to be obtained simultaneously. Considering the elongated body shape of shrimp, proportional padding was performed before resizing to reduce morphological distortion. This preprocessing pipeline ensures that all models are evaluated under consistent input conditions.

4.3 Evaluation Metrics

Different evaluation metrics were used for the single-label and dual-label recognition tasks. For the BD Fish & Shrimp Disease Dataset and TigerShrimpBD, Accuracy, Recall, and F1-score were used to evaluate classification performance:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{51}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{52}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{53}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \tag{54}$$

Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

For the ShrimpMTD dual-label dataset, Species Accuracy and Disease Accuracy were calculated to evaluate species recognition and disease recognition separately. To further measure the overall performance of the species–disease dual-label task, Weighted Joint Accuracy (WJA) was introduced:

$$\text{WJA} = 0.3 \times \text{Acc}_s + 0.7 \times \text{Acc}_d \tag{55}$$

Here, Species Accuracy (Acc_s) denotes the accuracy of shrimp species recognition, and Disease Accuracy (Acc_d) denotes the accuracy of disease recognition. Since

disease recognition is the primary objective of this study, a higher weight is assigned to Disease Accuracy. WJA is not intended to replace Species Accuracy or Disease Accuracy, but serves as a complementary metric for evaluating the overall performance of the dual-label task.

In addition, the parameter count (Params) and the number of floating-point operations (FLOPs) were used to evaluate model complexity. Params reflects the model size, while FLOPs measures computational cost. These metrics allow recognition performance and lightweight efficiency to be analyzed jointly.

4.4 Comparative Model Settings

To verify the effectiveness of MSLG-Net, several representative models were selected for comparison, including the baseline backbone, general lightweight networks, a higher-complexity model from the EfficientNet series (Tan and Le 2019; Liu 2020), and shrimp disease recognition-related improved models.

EfficientNet-Lite0 was used as the baseline backbone to evaluate the improvement brought by the proposed modules. GhostNet (Han et al. 2020), MobileNetV3-Small and MobileNetV4-Small (Howard et al. 2019; Qin et al. 2025) were selected as representative lightweight models, while EfficientNet-B0 was included to compare MSLG-Net with a higher-complexity model from the same series. In addition, FeatherNetX (Sharma et al. 2026), SDNet (Varma and Krishna 2022), DICNN (Ramachandran et al. 2023), E-DarkNet53 and ConvNeXt-CBAM (Lou et al. 2026; Le et al. 2025) were used as shrimp disease recognition-related improved models, covering lightweight design, attention mechanisms, multi-scale convolution, and feature fusion strategies.

All comparative models were trained and evaluated under the same data splits, input size, training strategy, and main hyperparameter settings. The evaluation metrics followed those defined in Section 4.3. The selected comparative models and their purposes are summarized in Table 2.

Table 2 Comparative models and experimental purposes

Model category	Comparative models	Experimental purpose
Baseline backbone model	EfficientNet-Lite0	To evaluate the improvement over the original backbone
General lightweight models	MobileNetV3-Small, MobileNetV4-Small, GhostNet	To compare different lightweight architectures
Same-series model	EfficientNet-B0	To compare with a higher-complexity EfficientNet model
Improved models	FeatherNetX, ConvNeXt-CBAM, DICNN, E-DarkNet53, SDNet	To compare related attention, fusion, multi-scale, and lightweight designs
Proposed model	MSLG-Net	To evaluate the proposed local–global fusion framework

5 Experimental Results and Analysis

5.1 Comparative Experimental Results

MSLG-Net was evaluated on the BD Fish & Shrimp Disease Dataset and TigerShrimpBD for single-label disease classification. The results are shown in Table 3 and Table 4.

On the BD Fish & Shrimp Disease Dataset, MSLG-Net achieved an Accuracy of 0.9551, a Recall of 0.9225, and an F1-score of 0.9228, improving over EfficientNet-Lite0 by 0.0275, 0.0394, and 0.0396, respectively. Although E-DarkNet53 and DICNN obtained higher accuracy, they required much higher computational costs. In contrast, MSLG-Net used only 2.999M parameters and 0.733G FLOPs, showing a better balance between performance and complexity.

On TigerShrimpBD, MSLG-Net achieved an Accuracy of 0.9832, a Recall of 0.9836, and an F1-score of 0.9834, with Accuracy and F1-score improved by 0.0335 and 0.0329 over EfficientNet-Lite0. These results indicate that the proposed local lesion perception, global morphology modeling, and local–global fusion mechanisms improve the feature representation ability of the lightweight backbone across different shrimp disease datasets.

Table 3 Classification performance comparison on the BD Fish & Shrimp Disease Dataset. Numerical values for MSLG-Net are shown in bold

Model	Params (M)	FLOPs (G)	Accuracy	Recall	F1
EfficientNet-Lite0	3.385	0.768	0.9276 ± 0.0055	0.8831 ± 0.0075	0.8832 ± 0.0065
EfficientNet-B0	4.022	0.769	0.9506 ± 0.0056	0.9160 ± 0.0110	0.9164 ± 0.0111
MobileNetV3-Small	1.529	0.111	0.9360 ± 0.0063	0.8967 ± 0.0123	0.8960 ± 0.0117
MobileNetV4-Small	2.507	0.369	0.9287 ± 0.0032	0.8873 ± 0.0073	0.8882 ± 0.0076
GhostNet	3.916	0.280	0.9439 ± 0.0040	0.9052 ± 0.0056	0.9047 ± 0.0056
E-DarkNet53	40.626	14.936	0.9697 ± 0.0024	0.9467 ± 0.0045	0.9461 ± 0.0046
DICNN	7.264	11.041	0.9686 ± 0.0021	0.9447 ± 0.0043	0.9441 ± 0.0044
FeatherNetX	0.740	0.347	0.9540 ± 0.0081	0.9217 ± 0.0109	0.9207 ± 0.0113
ConvNeXt-CBAM	27.926	0.644	0.9506 ± 0.0057	0.9217 ± 0.0094	0.9219 ± 0.0088
SDNet	25.711	0.190	0.9383 ± 0.0044	0.9200 ± 0.0071	0.9222 ± 0.0070
MSLG-Net	**2.999**	**0.733**	$\mathbf{0.9551 \pm 0.0029}$	$\mathbf{0.9225 \pm 0.0057}$	$\mathbf{0.9228 \pm 0.0056}$

Table 4 Classification performance comparison on TigerShrimpBD. Numerical values for MSLG-Net are shown in bold

Model	Params (M)	FLOPs (G)	Accuracy	Recall	F1
EfficientNet-Lite0	3.376	0.768	0.9497 ± 0.0243	0.9510 ± 0.0242	0.9505 ± 0.0238
EfficientNet-B0	4.013	0.769	0.9541 ± 0.0158	0.9551 ± 0.0154	0.9545 ± 0.0155
MobileNetV3-Small	1.522	0.111	0.9375 ± 0.0597	0.9381 ± 0.0608	0.9384 ± 0.0591
MobileNetV4-Small	2.498	0.369	0.9585 ± 0.0103	0.9611 ± 0.0096	0.9593 ± 0.0100
GhostNet	3.907	0.280	0.9521 ± 0.0172	0.9533 ± 0.0164	0.9527 ± 0.0165
E-DarkNet53	40.619	14.936	0.9855 ± 0.0037	0.9863 ± 0.0035	0.9858 ± 0.0036
DICNN	7.257	11.041	0.9716 ± 0.0046	0.9723 ± 0.0036	0.9718 ± 0.0042
FeatherNetX	0.739	0.347	0.9837 ± 0.0057	0.9843 ± 0.0056	0.9839 ± 0.0057
ConvNeXt-CBAM	27.920	0.644	0.9762 ± 0.0041	0.9768 ± 0.0039	0.9767 ± 0.0040
SDNet	25.710	0.190	0.9524 ± 0.0060	0.9551 ± 0.0052	0.9532 ± 0.0060
MSLG-Net	**2.998**	**0.733**	$\mathbf{0.9832 \pm 0.0020}$	$\mathbf{0.9836 \pm 0.0023}$	$\mathbf{0.9834 \pm 0.0020}$

5.2 Ablation Study Results

To evaluate the contribution of each module, ablation experiments were conducted on the ShrimpMTD dataset, as shown in Table 5. The baseline model without LLP, GMP, and LGGF achieved a WJA of 0.7958 and a Disease Accuracy of 0.7490. Adding LLP increased WJA to 0.8225 and Disease Accuracy to 0.7820, indicating that local lesion perception improves disease-related feature discrimination. Adding GMP increased WJA to 0.8166 and Species Accuracy to 0.9272, suggesting that global morphology modeling contributes to species and disease representation.

When LLP and GMP were used together, WJA and Disease Accuracy further increased to 0.8381 and 0.8022, respectively. After introducing LGGF, the complete MSLG-Net achieved the best overall performance, with a WJA of 0.8470, a Species Accuracy of 0.9343, and a Disease Accuracy of 0.8096. These results demonstrate that local lesion features and global morphological features are complementary, and that adaptive local–global fusion further improves dual-label recognition performance.

The Params values in Table 5 represent the actual parameter counts of complete models under different ablation settings, rather than the simple sum of individual module parameters. When LLP or GMP is disabled, a 1×1 convolutional adaptation layer is used for feature alignment; when LGGF is disabled, parameter-free average fusion is adopted. Therefore, different feature projection and fusion paths may lead to slight non-monotonic changes in parameter count. All ablation settings use the same input size, output task, and statistical method to ensure fair comparison.

Table 5 Ablation study results of MSLG-Net. Numerical values for the complete MSLG-Net are shown in bold

Experimental setting	LLP	GMP	LGGF	WJA	Species Acc	Disease Acc	Params (M)
Baseline	×	×	×	0.7958 ± 0.0194	0.9050 ± 0.0122	0.7490 ± 0.0225	3.377
Baseline+LLP	✓	×	×	0.8225 ± 0.0237	0.9170 ± 0.0175	0.7820 ± 0.0253	3.128
Baseline+GMP	×	✓	×	0.8166 ± 0.0264	0.9272 ± 0.0143	0.7692 ± 0.0316	3.038
Baseline+LLP+GMP	✓	✓	×	0.8381 ± 0.0038	0.9219 ± 0.0137	0.8022 ± 0.0060	3.711
MSLG-Net	✓	✓	✓	$\mathbf{0.8470 \pm 0.0214}$	$\mathbf{0.9343 \pm 0.0169}$	$\mathbf{0.8096 \pm 0.0233}$	**3.020**

5.3 Dual-Label Recognition Results

The ShrimpMTD dataset requires simultaneous recognition of shrimp species and disease type, making it more challenging than single-label classification. As shown in Table 6, MSLG-Net achieved a WJA of 0.8470, a Species Accuracy of 0.9343, and a Disease Accuracy of 0.8096. Compared with EfficientNet-Lite0, these metrics improved by 0.0886, 0.0587, and 0.1015, respectively, indicating that the proposed local–global feature modeling is effective for the dual-label task.

Although E-DarkNet53 and DICNN obtained slightly higher WJA values of 0.8821 and 0.8561, respectively, they required much higher computational costs. E-DarkNet53 used 41.002M parameters and 7.515G FLOPs, while DICNN used 7.617M parameters and 5.522G FLOPs. In contrast, MSLG-Net required only 3.020M parameters and 0.393G FLOPs, achieving a more favorable balance between dual-label recognition performance and lightweight complexity.

It should be noted that the Params and FLOPs values in Table 6 are calculated under the ShrimpMTD dual-label setting, where all models are equipped with multi-task output heads for species, disease, and joint category prediction. Therefore, these values should not be directly compared with those in the single-label task tables. Model complexity is mainly compared within the same dataset and task setting.

Table 6 Comparison of dual-label recognition results on ShrimpMTD. Numerical values for MSLG-Net are shown in bold

Model	Params (M)	FLOPs (G)	WJA	Species Acc	Disease Acc
EfficientNet-Lite0	3.395	0.366	0.7584 ± 0.0065	0.8756 ± 0.0103	0.7081 ± 0.0135
EfficientNet-B0	4.032	0.385	0.7952 ± 0.0119	0.9308 ± 0.0046	0.7371 ± 0.0152
MobileNetV3-Small	1.537	0.048	0.7214 ± 0.0125	0.8676 ± 0.0045	0.6587 ± 0.0197
MobileNetV4-Small	2.517	0.185	0.7515 ± 0.0041	0.8900 ± 0.0078	0.6922 ± 0.0060
GhostNet	3.926	0.154	0.8133 ± 0.0143	0.9281 ± 0.0122	0.7641 ± 0.0162
E-DarkNet53	41.002	7.515	0.8821 ± 0.0167	0.9549 ± 0.0100	0.8507 ± 0.0196
DICNN	7.617	5.522	0.8561 ± 0.0204	0.9523 ± 0.0046	0.8149 ± 0.0272
FeatherNetX	0.742	0.188	0.8212 ± 0.0142	0.9139 ± 0.0068	0.7815 ± 0.0176
ConvNeXt-CBAM	27.932	0.332	0.7680 ± 0.0152	0.8585 ± 0.0183	0.7292 ± 0.0138
SDNet	25.712	0.094	0.7361 ± 0.0210	0.8186 ± 0.0260	0.7007 ± 0.0190
MSLG-Net	**3.020**	**0.393**	$\mathbf{0.8470 \pm 0.0214}$	$\mathbf{0.9343 \pm 0.0169}$	$\mathbf{0.8096 \pm 0.0233}$

5.4 Visualization Analysis

To further investigate the decision-making behavior of MSLG-Net, confusion matrices and gradient-weighted class activation mapping (Grad-CAM) visualizations are employed. The confusion matrices illustrate the distribution of correct and incorrect predictions, while Grad-CAM highlights the image regions contributing to model decisions, where warmer areas indicate stronger responses.

As shown in Figs. 6–9, MSLG-Net achieved highly concentrated diagonal distributions on both public datasets and the ShrimpMTD dual-label task, indicating reliable classification capability. For ShrimpMTD, species recognition showed relatively higher stability than disease recognition because species-related morphological characteristics are generally more consistent, whereas disease categories often exhibit overlapping visual symptoms. The remaining errors mainly occurred among visually similar disease categories, such as black gill disease, vibriosis, and abnormal body color conditions. These misclassifications reveal that fine-grained disease recognition remains challenging due to subtle differences in color, texture, and lesion appearance.

Figs. 10–12 present Grad-CAM visualization comparisons among different models. Compared with baseline lightweight networks, MSLG-Net produces more compact and biologically meaningful activation regions. For diseased shrimp samples, the model mainly focuses on disease-related areas, including body surface abnormalities, gill regions, cephalothorax regions, and white-spot-like lesions. In contrast, some lightweight models exhibit scattered responses or excessive attention outside the lesion regions, indicating insufficient sensitivity to subtle pathological features. For healthy samples, MSLG-Net mainly activates the overall shrimp contour rather than isolated abnormal regions, suggesting that the model can distinguish normal morphological patterns from disease-related changes.

Overall, the visualization results indicate that MSLG-Net can focus on biologically meaningful and disease-related regions. This observation is consistent with the comparative and ablation results, further supporting the effectiveness of the proposed local–global feature fusion design.

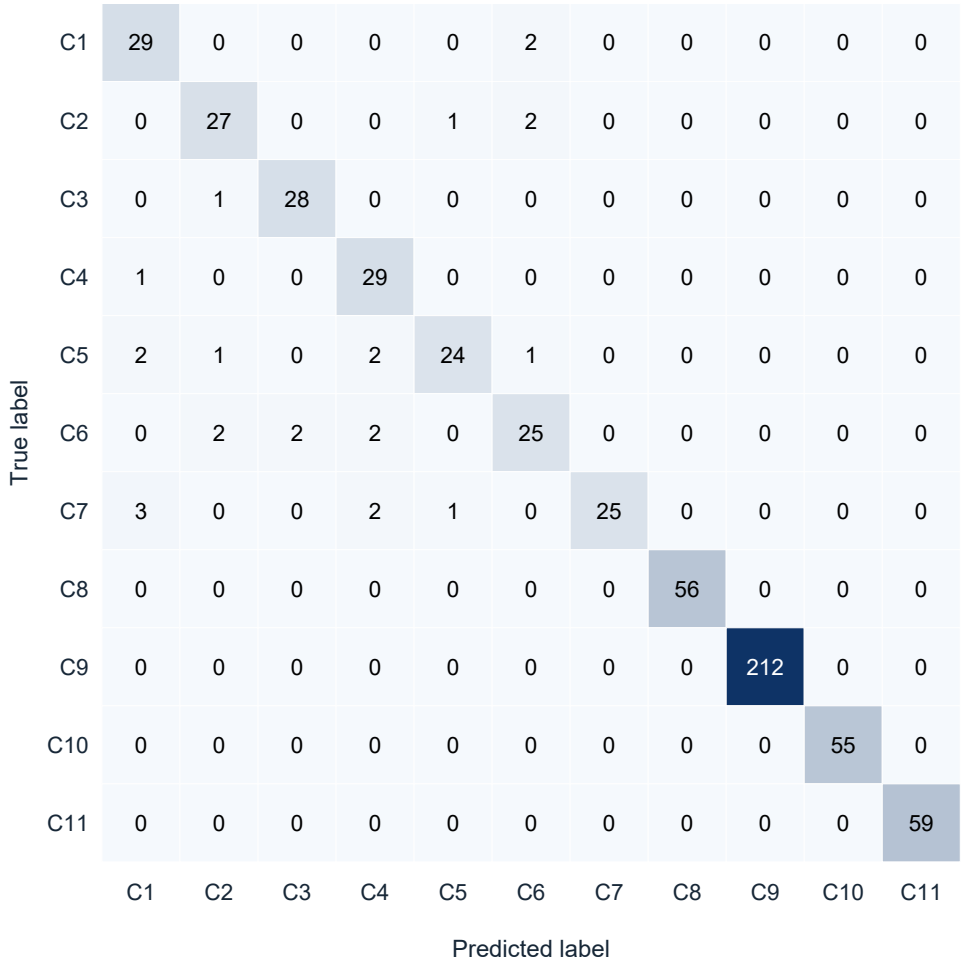


Fig. 6 Confusion matrix on the BD Fish & Shrimp Disease Dataset. Entries are sample counts; rows indicate true classes and columns indicate predicted classes. The class codes C1–C11 are defined below the matrix

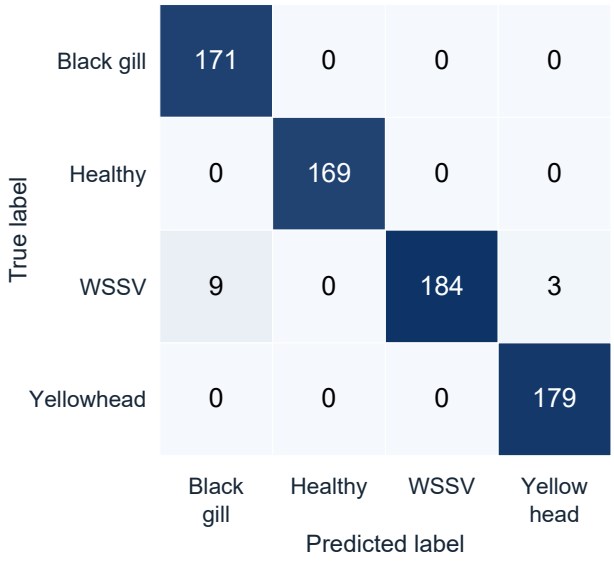


Fig. 7 Confusion matrix on TigerShrimpBD. Entries are sample counts; rows indicate true classes and columns indicate predicted classes

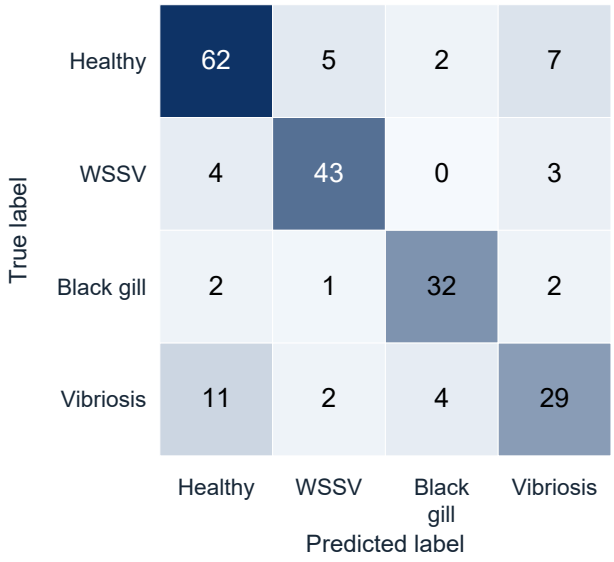
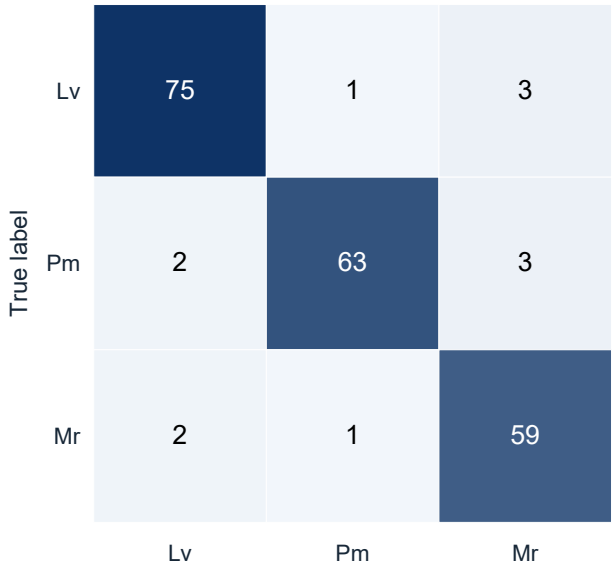


Fig. 8 Disease confusion matrix on ShrimpMTD. Entries are sample counts; rows indicate true classes and columns indicate predicted classes



Lv: Litopenaeus vannamei
Pm: Penaeus monodon
Mr: Macrobrachium rosenbergii

Fig. 9 Species confusion matrix on ShrimpMTD. Entries are sample counts; rows indicate true classes and columns indicate predicted classes. Species abbreviations are defined below the matrix

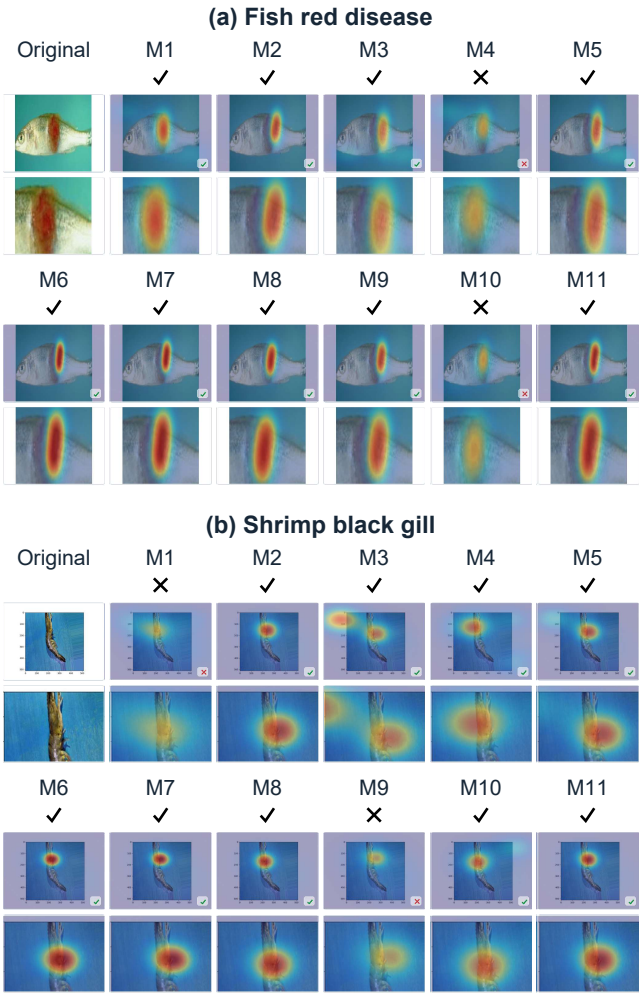


Fig. 10 Grad-CAM comparisons on the BD Fish & Shrimp Disease Dataset: (a) fish red disease; (b) shrimp black gill. Each sample has two blocks of six columns, with overview and detail views in the upper and lower rows of each block. Model codes are defined in the legend; check marks and crosses denote correct and incorrect predictions, respectively

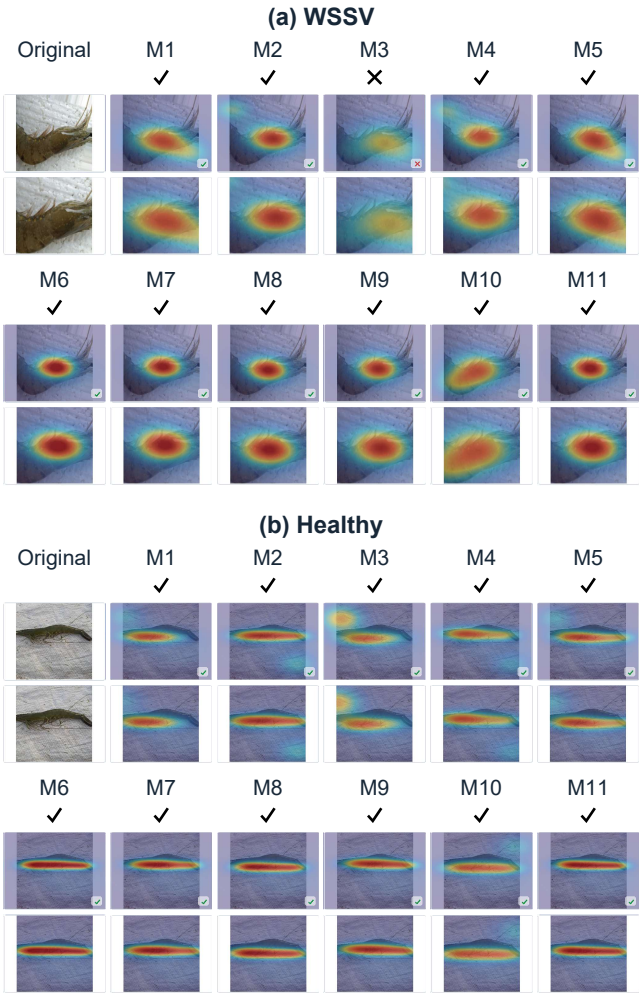


Fig. 11 Grad-CAM comparisons on TigerShrimpBD: (a) WSSV; (b) healthy. Each sample has two blocks of six columns, with overview and detail views in the upper and lower rows of each block. Model codes are defined in the legend; check marks and crosses denote correct and incorrect predictions, respectively

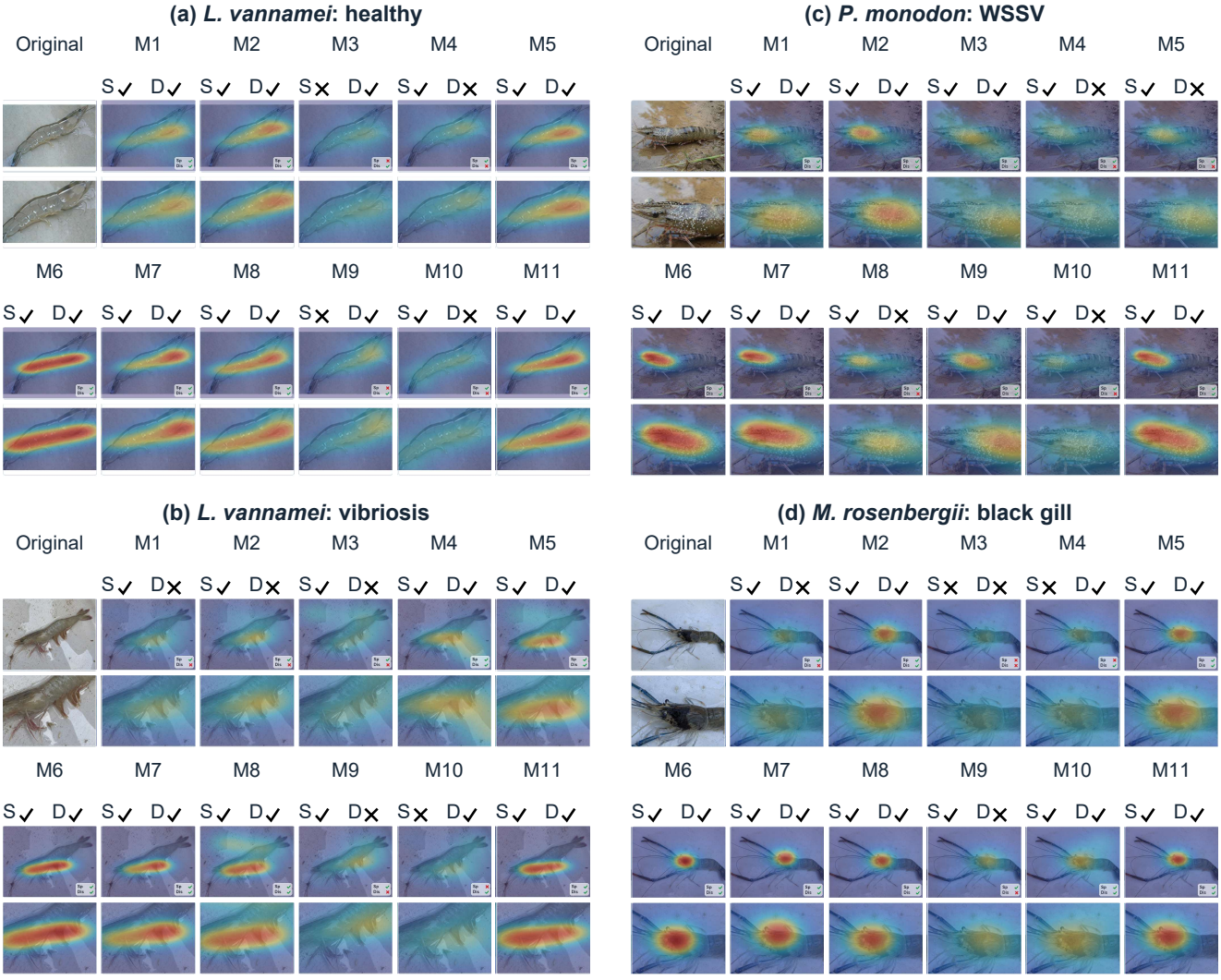


Fig. 12 Grad-CAM comparisons on ShrimpMTD: (a) healthy *L. vannamei*; (b) *L. vannamei* with vibriosis. Each sample has two blocks of six columns, with overview and detail views in the upper and lower rows of each block. Model codes are defined in the legend. S and D denote species and disease predictions; check marks and crosses denote correct and incorrect predictions, respectively

Fig. 12 (continued) (c) *P. monodon* with WSSV; (d) *M. rosenbergii* with black gill. The column, row, and prediction-marker conventions are the same as in panels (a) and (b)

6 Conclusion

This study proposed MSLG-Net, a lightweight local–global feature fusion network for shrimp disease recognition. Experiments on two public single-label datasets and the self-constructed ShrimpMTD dual-label dataset showed that the proposed model improves the EfficientNet-Lite0 baseline while maintaining moderate computational cost. Ablation and visualization results further verified the effectiveness of local lesion perception, global morphology modeling, and adaptive feature fusion. However, the current dataset scale remains limited, and visually similar disease categories are still difficult to distinguish. Future work will focus on expanding the dataset, improving annotation reliability, and optimizing the model for mobile and edge-device deployment.

References

Ahmed SI, Farid DM (2025) TigerShrimpBD: a tiger shrimp image dataset. Mendeley Data, https://doi.org/10.17632/9dj4sk5d55.1, version 1

Aung T, Vanichviriyakit R, Chayantrakom K, et al (2025) CNN-based identification of pathogens of concern in shrimp. Animals 15:3194. https://doi.org/10.3390/ani15213194

de Souza Valente C, Wan AHL (2021) Vibrio and major commercially important vibriosis diseases in decapod crustaceans. Journal of Invertebrate Pathology 181:107527. https://doi.org/10.1016/j.jip.2020.107527

Fei Y, Wang G, Liu F, et al (2025) Lightweight shrimp disease detection research based on YOLOv8n. arXiv preprint arXiv:2507.02354, https://doi.org/10.48550/arXiv.2507.02354

Frischer ME, Landers SC, Walker AN, et al (2022) Black gill in marine decapod crustaceans: a review. Reviews in Fisheries Science & Aquaculture 30:498–519. https://doi.org/10.1080/23308249.2022.2047153

Han K, Wang Y, Tian Q, et al (2020) GhostNet: more features from cheap operations. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 1577–1586, https://doi.org/10.1109/CVPR42600.2020.00165

Howard A, Sandler M, Chu G, et al (2019) Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, pp 1314–1324, https://doi.org/10.1109/ICCV.2019.00140

Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 7132–7141, https://doi.org/10.1109/CVPR.2018.00745

Islam MM, Rasel AASR (2024) ShrimpDiseaseImageBD: an image dataset for computer vision-based detection of shrimp diseases in Bangladesh. Mendeley Data, https://doi.org/10.17632/jhrtdj9txm.1, version 1

Jone IH, Masud DMRB, Sarker P, et al (2025) ShrimpXNet: a transfer learning framework for shrimp disease classification with augmented regularization, adversarial training, and explainable AI. In: 2025 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). IEEE, pp 1–6, https://doi.org/10.1109/3ICT68299.2025.11442114

Le TD, Huynh KT, Pham DD, et al (2025) Edge-IoT ConvNeXt-CBAM system for real-time classification of shrimp health conditions. In: 2025 International Conference on Advanced Technologies for Communications (ATC). IEEE, pp 1–6, https://doi.org/10.1109/ATC67618.2025.11268567

Liu R (2020) Higher accuracy on vision models with EfficientNet-Lite. TensorFlow Blog, URL https://blog.tensorflow.org/2020/03/higher-accuracy-on-vision-models-with-efficientnet-lite.html, accessed 3 Sep 2026

Lou P, Zhao Y, Ma S, et al (2026) An assessment method for the severity of shrimp red-leg disease based on E-DarkNet53 and multichannel feature fusion deep learning network. Smart Agricultural Technology 14:101855. https://doi.org/10.1016/j.atech.2026.101855

Munro J, Owens L (2007) Yellow head-like viruses affecting the penaeid aquaculture industry: a review. Aquaculture Research 38:893–908. https://doi.org/10.1111/j.1365-2109.2007.01735.x

Qin D, Leichner C, Delakis M, et al (2025) MobileNetV4: universal models for the mobile ecosystem. In: Leonardis A, Ricci E, Roth S, et al (eds) Computer Vision – ECCV 2024, Lecture Notes in Computer Science, vol 15098. Springer, Cham, pp 78–96, https://doi.org/10.1007/978-3-031-73661-2_5

Ramachandran L, Mohan V, Senthilkumar S, et al (2023) Early detection and identification of white spot syndrome in shrimp using an improved deep convolutional neural network. Journal of Intelligent & Fuzzy Systems 45:6429–6440. https://doi.org/10.3233/JIFS-232687

Sánchez-Paz A (2010) White spot syndrome virus: an overview on an emergent concern. Veterinary Research 41:43. https://doi.org/10.1051/vetres/2010015

Sandler M, Howard A, Zhu M, et al (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 4510–4520, https://doi.org/10.1109/CVPR.2018.00474

Saon SC (2025) BD Fish & Shrimp Disease Dataset. Hugging Face, URL https://huggingface.co/datasets/Saon110/bd-fish-disease-dataset, accessed 3 Sep 2026

Sharma S, Rumahorbo PS, Kondo S, et al (2026) A lightweight deep learning architecture for automatic shrimp disease classification. Scientific Reports 16:13837. https://doi.org/10.1038/s41598-026-44195-z

Tamut H, Ghosh R, Gosh K, et al (2025) Enhancing disease detection in the aquaculture sector using convolutional neural networks analysis. Aquaculture Journal 5:6. https://doi.org/10.3390/aquacj5010006

Tan M, Le Q (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 97. PMLR, pp 6105–6114, URL https://proceedings.mlr.press/v97/tan19a.html

Varma GT, Krishna AS (2022) SDNet: integrated unsupervised learning with DLCNN for shrimp disease detection and classification. In: 2022 IEEE International Conference on Data Science and Information System (ICDSIS). IEEE, pp 1–6, https://doi.org/10.1109/ICDSIS55133.2022.9915812

Wang Q, Wu B, Zhu P, et al (2020) ECA-Net: efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 11531–11539, https://doi.org/10.1109/CVPR42600.2020.01155

Woo S, Park J, Lee JY, et al (2018) CBAM: convolutional block attention module. In: Ferrari V, Hebert M, Sminchisescu C, et al (eds) Computer Vision – ECCV 2018, Lecture Notes in Computer Science, vol 11211. Springer, Cham, pp 3–19, https://doi.org/10.1007/978-3-030-01234-2_1

Wu AQ, Li KL, Song ZY, et al (2025) Deep learning for sustainable aquaculture: opportunities and challenges. Sustainability 17:5084. https://doi.org/10.3390/su17115084

Xu W, Wan Y, Zhao W (2025) ELA: efficient location attention for deep convolution neural networks. Journal of Real-Time Image Processing 22:140. https://doi.org/10.1007/s11554-025-01719-6

Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122, https://doi.org/10.48550/arXiv.1511.07122

Statements and Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors declare no competing interests.

Ethics Approval

Not applicable.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Data Availability

The BD Fish & Shrimp Disease Dataset (Saon 2025) and TigerShrimpBD (Ahmed and Farid 2025) used in this study are publicly available from their respective repositories. The self-constructed ShrimpMTD dataset is hosted in a private Hugging Face repository (https://huggingface.co/datasets/zhangsanhang/ShrimpMTD). Because of the substantial effort involved in collecting ShrimpMTD, the authors currently provide the dataset through controlled access. Requests for research access should be directed to the corresponding author and are subject to author approval.

Code Availability

The source code of MSLG-Net is publicly available at: https://github.com/zhang08206-dev/MSLG-Net

Authors' Contributions

Yuhang Mao: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Kaiyan Zhu: Supervision, Writing – review and editing. Meini Wang: Data curation, Investigation. Shikai Zhang: Validation, Formal analysis. Yuanbo Dai: Writing – review and editing, Technical support. All authors read and approved the final manuscript.