

## RESEARCH ARTICLE

# DAMNet: Dual Attention Mechanism Deep Neural Network for Underwater Biological Image Classification

PEIXIN QU<sup>1</sup>, TENGFEI LI<sup>1</sup>, LING ZHOU<sup>1</sup>, SONGLIN JIN<sup>1</sup>, ZHENG LIANG<sup>2</sup>, WENYI ZHAO<sup>3</sup>, AND WEIDONG ZHANG<sup>1</sup>, (Member, IEEE)

<sup>1</sup>School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China

<sup>2</sup>School of Internet, Anhui University, Hefei, Anhui 230039, China

<sup>3</sup>School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Weidong Zhang (zwd\_wd@163.com)

This work was supported in part by the Natural Science Foundation of Henan Province under Grant 212300410345, and in part by the Key Specialized Research and Development Program of Science and Technology in Henan Province under Grant 212102110298 and Grant 222102210171.

**ABSTRACT** Due to the complex background and biodiversity of underwater biological images makes the identification of marine organisms difficult. To solve these above problems, we propose a dual attention mechanism deep neural network for underwater biological image classification (DAMNet). Firstly, the proposed DAMNet uses multi-stage stacking to suppress the complex underwater background, and the multiple stacking can reduce the number of parameters of the model and improve the generalization ability. Secondly, the dual attention mechanism module is combined with the improved reverse residual bottleneck based on deep convolution to extract the feature information of underwater biological images from space and channel aspects to obtain better discrimination and feature extraction capability. Finally, the gravity optimizer is selected to update the model weights, and the exponential translation can improve the model's convergence speed and learning rate. Extensive experiments on a dataset consisting of seven types of underwater biological images demonstrate that the DAMNet model has higher learning ability and robustness compared to the state-of-the-art methods. Our DAMNet model achieves 96.93% classification accuracy in all categories, which is at least a 2 percentage point improvement compared to other models.

**INDEX TERMS** Underwater images, attention mechanism, neural networks, reverse residuals, optimization algorithms.

## I. INTRODUCTION

Nowadays, underwater images have been increasingly studied in image enhancement, restoration, and classification [1]. Understanding marine organisms and abundance distribution is important for marine ecosystems, environmental monitoring, and fisheries [2], [3]. For underwater images, due to the scattering and absorption effects of light, suffer from image blur [4], low contrast [5], and blurred details [6], [7], [8], [9], [10], and the diversity of underwater image backgrounds is also an important influencing factor for classification.

The associate editor coordinating the review of this manuscript and approving it for publication was Fan Zhang<sup>1</sup>.

Therefore, classifying marine organisms is a very challenging and critical research topic.

Underwater biometric identification is gradually applied to aquacultures, such as automatic fishing and underwater survey, which is of great significance in fishery and seabed detection. There is much research on underwater image classification, but few effective underwater image recognition models exist. Image classification methods are still largely unsatisfactory for underwater image classification due to the poor performance of extracting salient features from images due to the diversity of underwater image backgrounds and image quality [11], [12], [13].

Underwater images have complex backgrounds and certain feature similarities among different species, so how to extract

valuable information from the complex images is the key to solving the problem. For the underwater species classification problem, this paper proposes a DAMNet model, which uses a dual-attention mechanism to extract rich image feature information and uses a gravity optimization algorithm to improve the accuracy of the model. Experiments show higher model training accuracy than common attention mechanisms and optimizers. **The main contributions of this paper are as follows.**

- In this paper, we propose a method for classifying images of underwater organisms and conduct many comparative experiments on images of different underwater organisms in different environments. Compared with other methods, our DAMNet model has better classification ability.
- We use multi-stage stacking method, to reduce the number of feature parameters by stacking modules and update the weight information through the extraction of underwater features and the change of parameter gradients. And the gravity optimization algorithm uses exponential moving averages, which can effectively improve the classification performance of the model.
- A self-attention mechanism is added to the model in this paper, and a classification module combining channel and spatial attention is used to make the model have better feature description ability. Then, further using residual network and global variable features, a better underwater image feature extraction method is proposed, which improves the classification accuracy and practical application value.

In what follows, section II briefly reviews the classification methods in underwater images. Section III presents the overall framework and methodology of the model. Section IV provides a comparative analysis of different models and ablation studies. Finally, Section V gives the conclusion and outlook of this paper.

## II. RELATED WORKS

In this section, we summarize the related work from two aspects, including machine learning methods and deep learning methods.

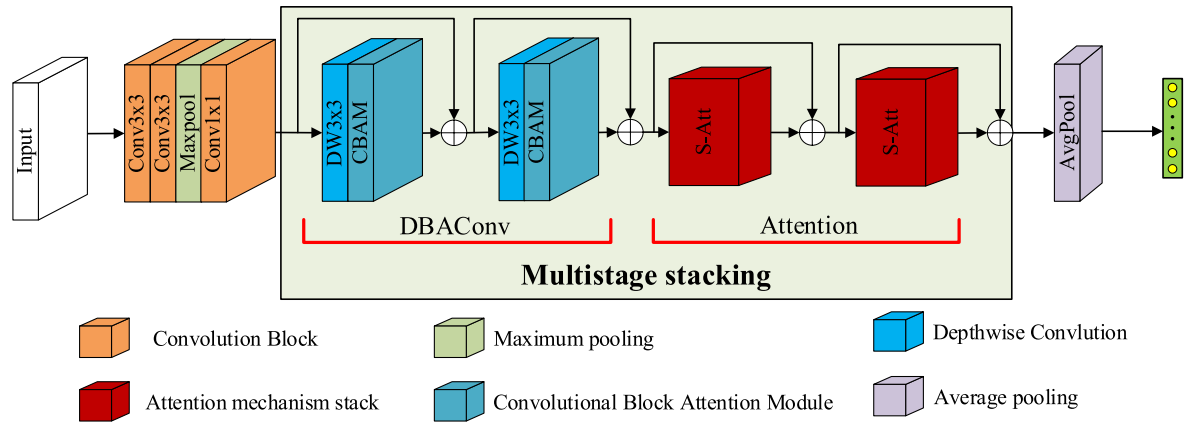
### A. MACHINE LEARNING METHODS

The classification method based on machine learning is semi-automatic extraction of image features, which requires manual definition and verification of functions. It uses algorithms to convert images into data to extract desired target features. Sun et al. [14] used a Support Vector Machine (SVM) as the final classification algorithm for the proposed low-resolution underwater image classification model. Iswari et al. [15] used K-nearest neighbors (KNN) as a classification algorithm for color summarization of fish images and achieved 91.36% classification accuracy. Deep and Dash [16] used KNN and SVM combined with convolutional neural networks (CNN) to classify underwater fish species and had

better results compared to algorithms. Zhang et al. [17] classified eight different fish species based on a genetic algorithm feature construction method and a histogram of orientation gradients, and their method obtained an average accuracy of 98.9%. Cheng et al. [18] showed that combining support vector machines for the classification of underwater plankton images can improve the accuracy and recall of the model, with the best result reaching 94%. Wang et al. [19] compared machine learning and deep learning and experimentally showed that machine learning has better results on small sample datasets. Salman et al. [20] combined machine learning processing methods to classify 15 and 10 underwater fish species and showed more than 90% accuracy. Khishe and Mosavi [21] compared an artificial neural network (ANN) trained chimpanzee optimization method with an iron motion algorithm and demonstrated the reliability of the algorithm. Almero et al. [22] used a hybrid model composed of hidden layers in an ANN and achieved an accuracy of 93.6% in fish detection. Machine learning has good fault tolerance and computational power, and convolutional neural networks have shorter training times and higher accuracy rates. Still, manually extracted features are not rich, have weak generalization ability, and are inefficient.

### B. DEEP LEARNING METHODS

The deep learning method to process images is to extract features, emphasizing the importance of feature learning automatically and feature extraction and transformation in multi-layer networks, so that deep learning has excellent feature expression capabilities. Mahmood et al. [23] used residual networks to extract new image features from different convolutional layers and combine them to obtain more compact and powerful deep features. Multi-channel convolution also has outstanding contributions in image classification, recovery and enhancement [24]. Xu et al. [25] improved the classification performance of deep convolutional neural networks in underwater images by optically transforming images and adding generative adversarial networks. Qi et al. [26] proposed a semantic region enhancement module that senses the degradation of different semantic regions from multiple scales and feeds them back to the global attention features extracted from their original scales to achieve enhancement of underwater images. Li et al. [27] proposed an underwater image enhancement network with multicolor spatial embedding guided by media transport, coupled with an attention mechanism where the most discriminative features extracted from multiple color spaces are adaptively integrated and highlighted. Xie et al. [28] normalized the total variance term and sparse prior knowledge sparse prior knowledge of the fuzzy kernel by varying the image resolution for fuzzy kernel estimation coarse to fine to avoid local minima, further validating its superiority relative to other state-of-the-art algorithms. Li et al. [29] investigated and studied deep learning methods applied to underwater image enhancement and evaluated the existing methods qualitatively and quantitatively.



**FIGURE 1. Flowchart of DAMNet model.** Our model consists of a dual attention mechanism module and a stacking module. In our approach, we pool and convolve image features, input them into the dual-attention mechanism module to extract feature information, and filter out the important features by stacking of modules. In the model, the redundancy of feature information is reduced at each stage using residual connections, and all convolutional layers are of size  $3 \times 3$  and  $1 \times 1$ .

Jaeger et al. [30] presented a Croatian fish dataset containing 12 fish images to perform fine visual classification of fish with an accuracy rate of 66.78%. Villon et al. [31] proposed a method to assist in recognition of fish species, using CNN to test that the recognition rate of fish reached 94.9%, which is higher than the human recognition rate. Balakrishnan et al. [32] used migration learning to classify underwater images, combined it with data augmentation, and achieved an accuracy of 86%. Aridos et al. [33] proposed a deep underwater image classification algorithm for turbid underwater images, which provides better classification accuracy. Fu et al. [34] use a graph CNN model to represent features in three aspects: local space, global space, and channel correlation, with better accuracy on real datasets. Yang et al. [35] proposed an initial attention network that outperformed other networks in distinguishing underwater images from images in non-underwater environments with 99.3% classification accuracy. Paraschiv et al. [36] used a lightweight CNN model to classify underwater fish, which improved the accuracy by 7% compared with the network model with many parameters. Mathur and Goel [37] proposed a migration learning based method for underwater image classification, which improved the classification accuracy by training only the last few layers of the network through migration learning and achieved 98.44% and 84.92% accuracy on large and small datasets, respectively. Jiang et al. [38] proposed an improved deep convolutional generative adversarial networks model to classify underwater objects, add data for small-sample objects, and combining the CNN models with lightweight neural network achieves a good trade-off between model complexity and classification accuracy. Deep learning methods have high generalization ability and robustness but large data volume and long training time.

### III. PROPOSED METHOD

We show the flowchart of our proposed DAMNet model in Fig 1. It includes the following steps: 1) multi-stage stacking,

2) dual-attention module, and 3) optimizer. Specifically, the first step uses multi-stage stacking to reduce the model's number of parameters and improve the model's generalization ability by combining and stacking between different modules and attention. The second step adds a dual-attention mechanism module to extract saliency information of images in space and channels to improve the model's classification accuracy. Finally, it uses a gravity optimization algorithm, a gradient-based optimization algorithm, to reduce the loss of the deep learning model and improve its training results.

#### A. MULTI-STAGE STACKING

For the training  $224 \times 224$  images, there are many pixels. After the convolution and attention are combined, if the relative attention is directly used, there will be problems such as a large amount of calculation, high cost, and slow speed due to many pixels. The amount of features is reduced without affecting the important features, and then relative attention is used.

The model adopts a multi-stage layout, and each stage uses max pooling to gradually reduce the space size and increase the number of channels, enabling residual networks to connect. The model is divided into five parts, containing  $S_0, S_1, S_2, S_3$  and  $S_4$ . The first part ( $S_0$ ) is used to extract features and increase the number of channels and is defined as:

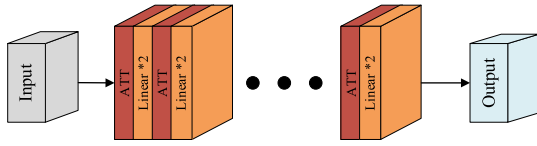
$$S_0 = \text{Conv}^3 \left( \text{Conv}^3 (x) \right), \quad (1)$$

where two  $3 \times 3$  convolutional layers ( $\text{Conv}^3$ ) and  $x$  are the input of the features.

The second ( $S_1$ ) and third ( $S_2$ ) parts combine convolution and self-attention to improve the model's accuracy, which can be expressed as:

$$y^M = \text{Conv}^1 \left( \text{Lin}^2 \left( \text{Avg} \left( \text{Conv}^3 \left( \text{Conv}^1 (x) \right) \right) \right) \right), \quad (2)$$

$$S_1, S_2 = y^M \left( y^M \left( \text{Conv}^1 \left( \text{Max} (x_1) \right) \right) \right), \quad (3)$$



**FIGURE 2.** Flowchart of attention stacking. Each attention stacking module is constructed by stacking  $n$  Transformer blocks.  $S_3$  consists of five Transformer blocks and  $S_4$  consists of two Transformer blocks, both with maximum pooling and  $1 \times 1$  convolution before stacking.

where  $S_1$  in  $x_1$  is  $S_0$ ,  $S_2$  in  $x_1$  is  $S_0 \oplus S_1$ ,  $\text{Conv}^1$  stands for  $1 \times 1$  convolution, Lin denotes fully connected, Avg denotes average pooling, and Max is maximum pooling.

The fourth ( $S_3$ ) and fifth ( $S_4$ ) parts are Transformer blocks, which purely use the attention mechanism to improve the efficiency of the data and the model by directly stacking the Transformer blocks concerning attention, and are defined as:

$$y^{\text{TF}} = \text{Lin}(\text{Lin}(\text{ATT}(x))), \quad (4)$$

$$S_3 = y^{\text{TF}*5}(\text{Conv}^1(\text{Max}^{*2}(x_2))), \quad (5)$$

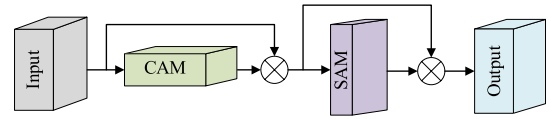
$$S_4 = y^{\text{TF}*2}(\text{Conv}^1(\text{Max}^{*2}(x_3))), \quad (6)$$

and  $y^{\text{TF}}$  can be seen as a Transformer block, both multi-headed attention and a location fully connected feedforward network (FFN), where  $*5$  or  $*2$  stands for stacking using this module 5 times or 2 times, and the stacking schematic is shown in Fig 2,  $x_2$  and  $x_3$  are  $S_1 \oplus S_2$  and  $S_2 \oplus S_3$ , respectively.

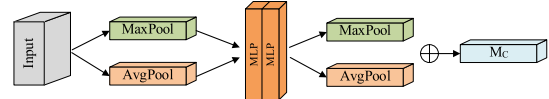
Transformer blocks are used later in the model to maximize the balance between efficiency and performance. Through the model's stage layout, many parameters are reduced to make the features reach a manageable level, which speeds up the model's training speed and helps improve the model's generalization ability when the training data is limited. When the data features are rich, the stacking of modules increases the model's capacity and improves the visual processing capability to a certain extent. High capacity and good generalization ability can effectively improve the classification performance of the model.

## B. DUAL ATTENTION MODULE

This paper introduces an attention mechanism to reduce the feature parameters extracted from the model. This strategy selectively focuses on regions where valuable features are closely related and thus ignores unimportant feature information. We use Depthwise Convolution and Convolutional Block Attention Module (CBAM) to form the DBAConv module, which can calculate and extract features without changing the channel and greatly reduce the number of parameters. The CBAM module structure diagram is shown in Fig 3. In Depthwise Convolution, one channel is only convolved by one convolution kernel, and the number of feature channels does not change, which can be expressed as each dimension in a predefined receptive field. The sum of



**FIGURE 3.** CBAM structure diagram. The CBAM module consists of two modules, spatial attention and channel attention, which infer feature maps along two dimensions, channel and spatial, and then multiply them with the original map for adaptive feature refinement, respectively.



**FIGURE 4.** CAM flow chart. The module compresses the spatial dimension by parallel global maxpool and global avgpool, then expands again to the original number of channels by perceptron MLP, sums these two outputs element by element, and finally uses Sigmoid activation to obtain the output of Channel Attention.

the weighted values is defined as:

$$y_n = \sum_{m \in \eta(n)} z_{n-m} \odot x_m, \quad (7)$$

$$M_D = \text{Conv}^{\text{group}}(\text{Conv}^1(x)), \quad (8)$$

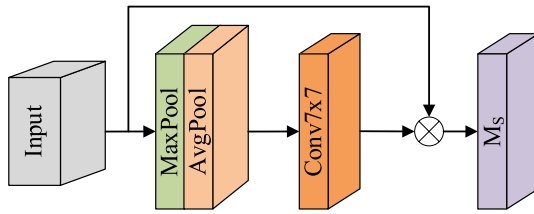
where  $x_n$  and  $y_n$  are the input and output of position  $n$ , respectively,  $\eta(n)$  is the local neighborhood of  $n$ , and  $z_{n-m}$  is the weight matrix of position  $(n - m)$ , respectively.

The attention mechanism CBAM module is a hybrid attention module consisting of channel attention (CAM) and spatial attention (SAM) modules, and the structure of the CAM is shown in Fig 4.  $\otimes$  is multiplied by the previous input,  $W, H$  represent the width and height of the feature map,  $C$  is the number of channels, and the input feature map  $F = W \times H \times C$ . For channel attention, the first is adaptive average pooling (Avg) and maximum adaptive pooling (Max), compressed into a feature map  $1 \times 1 \times C$ . Secondly, a two-layer MLP neural network is input, add two activation functions, and the output is  $1 \times 1 \times C$ . Get the same one-dimensional vector as the number of channels, learn the correlation between channels, and get the attention of the channels. Finally, the features computed by the module are subjected to a summation operation and a Sigmoid activation operation, resulting in  $(M_C)$  is expressed as:

$$M_C = \text{Sigmoid}(\text{MLP}(\text{Avg}(x)) + \text{MLP}(\text{Max}(x))), \quad (9)$$

where the neural network of the two-layer MLP,  $x$  is the input, the number of neurons in the first MLP is  $N/n$ , ( $n$  is the scaling rate, it is set to 16), and the number of neurons in the second MLP is  $N$ .

Unlike channel attention, spatial attention focuses on the effective information on feature map, SAM is shown in Fig 5. The results of channel attention output are first pooled equally and maximally to compute spatial attention to obtain two  $H \times W \times 1$  two-dimensional feature maps. Then the feature maps they produce are stitched together. Then splicing (Cat) into the spliced feature map of the feature map, the



**FIGURE 5. SAM flow chart.** The feature map output from Channel attention module is used as the input feature map of SAM module, the channel dimension is compressed by maxpool and avgpool, the two feature maps are stitched together, and after  $7 \times 7$  convolution operation, the dimension is reduced to one channel, and then the feature map of Spatial Attention is obtained by sigmoid activation, and finally the result is multiplied by the original map to obtain the original size feature map.

final spatial attention feature map ( $M_S$ ) is obtained by the  $7 \times 7$  convolution operation and the Sigmoid function, and it is defined as:

$$M_S = \text{Sigmoid} \left( \text{Conv}^7 \left( \text{Cat} \left( \text{Avg} (M_C); \text{Max} (M_C) \right) \right) \right). \quad (10)$$

The final element level is multiplied and the final CBAM attention module is defined as:

$$x' = M_C \otimes x, \quad (11)$$

$$M_F = M_S \otimes x', \quad (12)$$

where  $\otimes$  represents element-wise multiplication. Finally, the DBAConv module is defined as:

$$\text{DBA} = \text{Conv} (M_F (M_D (x))). \quad (13)$$

By extracting the feature information of the channel and space, multiplying and correcting the original input feature map, and generating the final feature map, which can suppress the complex underwater image noise information.

Although attention mechanisms have a larger capacity, they may generalize worse than CNN due to the lack of correct induction bias. The DBAConv module unifies and combines depthwise convolution and attention to make it more generalizable in classification. The stacking of Transformer blocks mainly establishes the global connection between features, while the DBAConv module establishes the local connection between features. The merging of global and local features can improve the model's performance.

Our final DAMNet mainly consists of a stack of DBA blocks and Transformer blocks, and the detailed building blocks and parameter settings are shown in Table 1. Before the first DBA module, the image size is reduced to  $112 \times 112$  and the number of channels input to the image is increased from two convolutional layers to 64. After the DBA module, the size of the feature map is reduced and the number of channels of the feature map is increased. The final prediction of the model is obtained after averaging pooling and fully connected layers.

**TABLE 1. Network building blocks and parameter settings for DAMNet.**

| Type                      | Kernel-size/stride/or remarks | Input size                 |
|---------------------------|-------------------------------|----------------------------|
| Conv                      | $3 \times 3/2$                | $224 \times 224 \times 3$  |
| Conv                      | $3 \times 3/1$                | $112 \times 112 \times 64$ |
| DBA module                | As in figure3                 | $112 \times 112 \times 64$ |
| DBA module                | As in figure3                 | $56 \times 56 \times 96$   |
| DBA module                | As in figure3                 | $56 \times 56 \times 96$   |
| DBA module                | As in figure3                 | $28 \times 28 \times 192$  |
| DBA module                | As in figure3                 | $28 \times 28 \times 192$  |
| $S - ATT \times 5$ module | As in figure2                 | $28 \times 28 \times 192$  |
| $S - ATT \times 3$ module | As in figure2                 | $14 \times 14 \times 384$  |
| AvgPool                   | $7 \times 7/1$                | $7 \times 7 \times 768$    |
| Linear                    | logits                        | $1 \times 1 \times 768$    |
| Softmax                   | classifier                    | $1 \times 1 \times 7$      |

### C. OPTIMIZER

In this paper, we use Gravity optimizer, a gradient-based optimization algorithm that reduces the parameters of the deep learning model. The gravity optimization algorithm has three hyperparameters, namely the learning rate ( $l$ ), the initial step size ( $\alpha$ ), and the average moving parameter ( $\beta$ ). The recommended values are  $l=0.1$ ,  $\alpha=0.01$ , and  $\beta=0.9$ , respectively, and the learning rate is a common hyperparameter in deep learning. Every optimization algorithm takes gradient as input and gives steps to update parameters in output. Where  $N$  is a normal distribution, the mean is  $\gamma$ , the standard deviation is  $\sigma$ , and  $G$  the gradient of the objective function. Where  $v=N(\gamma, \sigma)$ ,  $\sigma=\alpha/l$ ,  $t$  are the number of update steps.  $\beta$  is a positive real number between 0 and 1, and values that are too large or too small affect the training results, the proposed substitution scheme is defined as:

$$o = (\beta t + 1)/(t + 2), \quad (14)$$

and for each weight matrix, the maximum gradient calculation formula is defined as:

$$m = 1/\max(\text{abs}(G)). \quad (15)$$

The main problem before applying the moving average is the initial trial delay of loss reduction. To solve this problem, the gradient terms  $n$ ,  $v_t$  are the velocity in the current update step, which are defined as:

$$n = G/(1 + (G/m)^2), \quad (16)$$

$$v_t = o v_{t-1} + (1 - o) n, \quad (17)$$

The final weight update is defined as:

$$W^i = W^i - l v_t. \quad (18)$$

Among them, the larger the value of the gradient coefficient, the larger the step size of the weight with the higher gradient. By increasing  $m$ , the gradient of a larger range can be linearly processed, and the weight with a larger gradient value will also take a larger update: step size, low validation loss, and high accuracy. The experimental results are shown in Section IV-B.

#### D. LOSS FUNCTION

The sample used in this study is 9,500 underwater images of seven types, and different types of images have similar backgrounds. In the feature extraction process, to improve the learning speed of the model in the complex underwater image background and speed up the convergence of the model, the Cross-Entropy (CET) loss function in classification loss is used in this paper.

In the DAMNet model, the formula of the CET loss function is defined as:

$$C = -\frac{1}{n} \sum_x [r \ln f + (1 - r) \ln(1 - f)], \quad (19)$$

where  $x$  is the sample,  $f$  is the predicted input,  $r$  is the actual label, and  $n$  is the total number of samples. The formula is the total loss function of  $n$  samples, as long as the  $n$  losses are superimposed. Using the Sigmoid for regression can solve the problem that the weight update is too slow.

Extracting similar features from different classes of underwater images in the feature extraction process is affected by the error in the output when the data converge. We use the CET loss function can avoid the problem of decreasing the learning rate. When convergence is faster, the gradient of the last layer of weights is proportional to the difference between the output and true values and is no longer related to the activation function. The backpropagation is multiplicative, so the whole weight matrix is updated faster.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

*Experimental Dataset:* This paper selects datasets including fish, turtles, sea urchins, sea cucumbers, corals, humans, and remains from OceanDark [39], RUIE [40], UIEB [41], UFO-120 [42], EUVP [43], and publicly available online and other underwater biological samples. We collected about 5,800 underwater images containing different degraded and enhanced images such as real underwater images, artificial low-light images, hyper-segmented images and underwater images acquired by different devices, expanded them to about 9,500 images by image filtering and rotation, and divided them into training and test sets with a ratio of 5:1, with the training set containing 7,984 images and the test set containing 1,600 images. The size of each training sample was scaled to  $224 \times 224$ , and images of some different species underwater are shown in Fig 6.

*Compared methods:* This paper adopts the traditional and classic neural networks are used to compare and analyze the network models proposed in this paper, such as AlexNet [44], VGG19 [45], GoogLeNet [46], and ResNet50 [47], and new network models EfficientNe [48], CoAtNet [49], RepVGG [50] and AlterNet [51].

*Evaluate metrics:*

1) Accuracy(ACC): The percentage of total samples with all correct predictions.

2) Precision(P): The proportion of correct predictions as true samples to the total number of predictions as true samples.

**TABLE 2. Optimizer comparison of experimental results. The best results in each case are in red, with the next best in blue.**

| Optimizer      | ACC(%)       | P(%)         | R(%)         | F1(%)        | Loss          |
|----------------|--------------|--------------|--------------|--------------|---------------|
| Adam           | 94.80        | 94.71        | 94.35        | 94.53        | 0.3082        |
| AdamW          | 94.62        | 94.29        | 94.46        | 94.37        | 0.3429        |
| SGD            | 92.37        | 91.93        | 91.78        | 91.86        | 0.3219        |
| RMSProp        | 94.87        | 94.52        | 94.63        | 94.57        | 0.4504        |
| <b>Gravity</b> | <b>95.81</b> | <b>95.51</b> | <b>95.54</b> | <b>95.52</b> | <b>0.2337</b> |

3) Recall(R): The proportion of correctly predicted as true samples to the total actual as true samples.

4) F1: It is the summed average of Precision and Recall,  $F1 = P * R * 2 / (P + R)$ .

#### A. EXPERIMENTAL CONFIGURATION

This paper uses the DAMNet model for feature extraction and learning. On the personal workstation, RTX 3090 GPU and NVIDIA Tesla V100s GPU server with 32G graphics card as the device of the model, and implemented using the PyTorch framework.

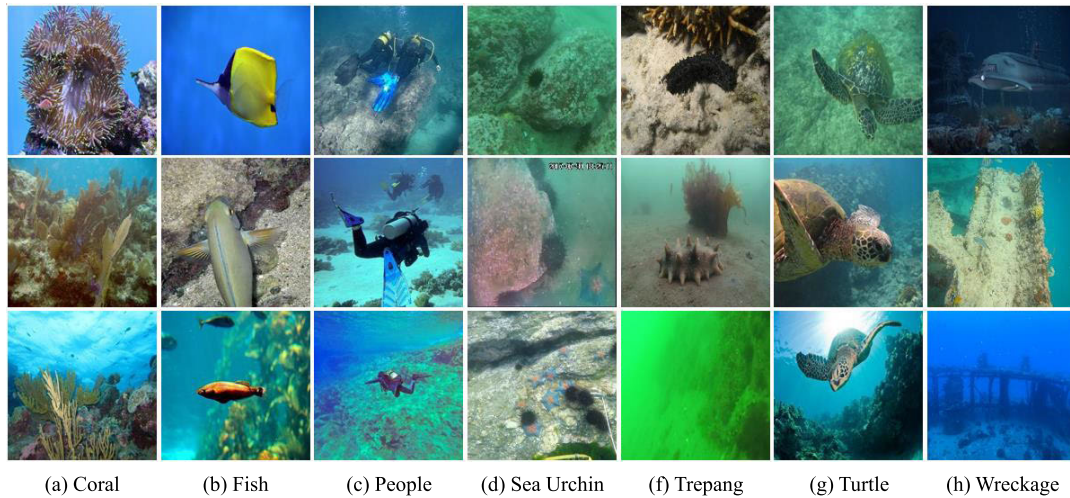
The original underwater creature images were uniformly converted to  $256 \times 256 \times 3$  and the input images were cropped to  $224 \times 224 \times 3$  by the center. The model was trained on 9,584 images, divided into a training set and a test set in a 5:1 ratio. The training set is used to train and optimize our DAMNet model, and the validation set is used to verify the validity of the model. We first initialize the network parameters using normalization and learn the model globally iteratively using the gravity optimizer, where we set the batch size to 16, the decay rate of the optimizer to  $\beta = 0.9$ , the base learning rate is set to 0.01, and the training ends with 200 iterations. In the following sections, the experimental results of classification evaluation, loss analysis, and ablation study are presented.

#### B. OPTIMIZER SELECTION

In this paper, the gravity optimization algorithm was selected as the optimizer of the model and compared with Adam, AdamW, SGD, and RMSProp optimizers. The experimental results are shown in Table 2 by training on the underwater biological image dataset. The results show that the gravity optimizer has lower verification loss, faster convergence, and higher accuracy. The RMSProp optimizer has a high accuracy rate, but the average loss is higher than other optimizers. Similar to the SGD optimizer, their average loss fluctuates greatly, and the convergence speed is slow. For different datasets, having a good optimizer to achieve faster weight updates and model convergence improves the model's classification accuracy and reduces the model's loss of parameters and training time.

#### C. CLASSIFICATION EVALUATION

As shown in Figure 6, this paper's training and test images are composed of seven types of biological images under different complex backgrounds and degradation conditions.



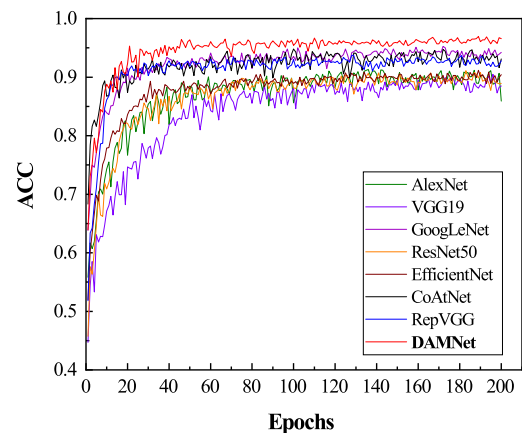
**FIGURE 6.** Image display of some underwater species. Our dataset contains seven different types of images of marine life, from left to right are samples of different kinds of underwater images. From top to bottom are samples of underwater images of the same species with different backgrounds.

The DAMNet model is used to classify the underwater biological image dataset, and the common network models are compared and tested. Such as AlexNet [44], VGG19 [45], GoogLeNet [46], and ResNet50 [47], and novel network models EfficientNe [48], CoAtNet [49], RepVGG [50] and AlterNet [51]. In Table 3, the classification accuracies of each class of underwater creature images are shown. The DAMNet model has a high accuracy in each class, but the accuracy of fish is lower, reaching only 91.84%, while the accuracy of all other species is around 97.5%.

The average training results of different models in all categories of the dataset are shown in Table 4, where we list the other models' accuracy, precision, recall, F1 value, and loss, verifying the learning and classification ability of the DAMNet model. The broken line comparison of training accuracy is shown in Fig 7.

Table 4 shows that the DAMNet model accurately classifies underwater images in complex and similar backgrounds. Compared with models that are purely deep convolutional networks or incorporate a single attention mechanism, the addition of a dual-attention model can extract richer image features, making the model complete in extracting object features, which can identify underwater creatures well and improve the classification accuracy of the model. The DAMNet model outperforms other network models in all metrics, with an accuracy rate of 96.93% and an average loss value of 0.1860. The accuracy is about 5 percentage points higher than that of models such as VGG19, EfficientNet, ResNet50, and AlexNet, respectively, and about 2 percentage points higher than that of models such as GoogLeNet, CoAtNet, RepVGG, and AlterNet.

Combined with Fig 7, it is found that the stacking of attention and the combination of convolution and attention make the DAMNet model converge faster, reach high accuracy in a short epoch, and smooth out compared with other



**FIGURE 7.** Accuracy curves of different model training processes. Our model reaches high precision and stabilizes at about 20 iterations, with fast convergence and high accuracy compared to other models.

models. The results in Fig 7 show that the DAMNet model has a good learning ability and convergence capacity to obtain satisfactory classification results for underwater biological images.

Combined Table 3 shows that the classification accuracy of fish is low in all the models with higher accuracy, and we found that the main reason is that the fish in the dataset is a large class containing different classes of fish, which makes the model have the low similarity of features in feature extraction and errors in classification.

#### D. LOSS ANALYSIS

In this paper, the data is increased by rotation, and the ratio of samples used for training and testing is 5:1, which reduces the risk of overfitting the model training. We initialize the weights and use a gravity optimization algorithm with a batch

**TABLE 3.** Experimental results for different categories of different models. The best results in each case are in red, with the next best in blue.

| Model/ACC(%)      | Coral        | Fish         | People       | Sea Urchin   | Trepang      | Turtle       | Wreckage     |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| AlexNet [44]      | 92.31        | 86.24        | 96.48        | 95.59        | 95.35        | 82.80        | 86.55        |
| VGG19 [45]        | 87.13        | 89.75        | 95.13        | 97.37        | 94.91        | 80.19        | 85.28        |
| GoogLeNet [46]    | <b>99.08</b> | 85.43        | 95.75        | <b>97.85</b> | <b>98.48</b> | <b>94.39</b> | 93.93        |
| ResNet50 [47]     | 91.32        | 82.83        | 94.42        | 97.48        | 94.95        | 84.79        | 87.73        |
| EfficientNet [48] | 88.60        | 88.40        | 94.63        | 95.71        | 92.96        | 87.19        | 87.29        |
| CoAtNet [49]      | 94.53        | 85.96        | <b>98.07</b> | 96.61        | 93.94        | 93.59        | <b>97.74</b> |
| RepVGG [50]       | 93.84        | <b>91.09</b> | 96.50        | 94.04        | 96.36        | 91.71        | 91.23        |
| AlterNet [51]     | 97.23        | 86.26        | 97.87        | 96.19        | 95.58        | 87.29        | 92.18        |
| <b>DAMNet</b>     | <b>97.28</b> | <b>91.84</b> | <b>98.08</b> | <b>97.86</b> | <b>97.01</b> | <b>97.55</b> | <b>97.84</b> |

**TABLE 4.** Experimental results of different models. The best results in each case are in red, with the next best in blue.

| Model             | ACC(%)       | P(%)         | R (%)        | F1(%)        | Loss          |
|-------------------|--------------|--------------|--------------|--------------|---------------|
| AlexNet [44]      | 91.30        | 90.58        | 90.76        | 90.67        | 0.5091        |
| VGG19 [45]        | 90.49        | 89.85        | 89.97        | 89.91        | 0.7887        |
| GoogLeNet [46]    | <b>95.24</b> | <b>95.21</b> | <b>94.99</b> | <b>95.10</b> | <b>0.2327</b> |
| ResNet50 [47]     | 91.11        | 90.61        | 90.50        | 90.55        | 0.4796        |
| EfficientNet [48] | 91.18        | 90.50        | 90.68        | 90.59        | 0.4441        |
| CoAtNet [49]      | 94.80        | 94.71        | 94.35        | 94.53        | 0.3082        |
| RepVGG [50]       | 93.74        | 93.58        | 93.68        | 93.63        | 0.2766        |
| AlterNet [51]     | 93.49        | 92.93        | 93.23        | 93.08        | 0.3338        |
| <b>DAMNet</b>     | <b>96.93</b> | <b>96.70</b> | <b>96.78</b> | <b>96.74</b> | <b>0.1860</b> |

**TABLE 5.** Experimental results of different attention mechanisms in the model. The best results in each case are in red, with the next best in blue.

| Module       | ACC(%)       | P(%)         | R(%)         | F1(%)        | Loss          |
|--------------|--------------|--------------|--------------|--------------|---------------|
| +SENet       | <b>94.49</b> | <b>94.14</b> | <b>94.31</b> | <b>94.22</b> | <b>0.3475</b> |
| +SKNet       | 94.12        | 93.63        | 93.86        | 93.44        | 0.3695        |
| <b>+CBAM</b> | <b>96.93</b> | <b>96.70</b> | <b>96.78</b> | <b>96.74</b> | <b>0.1860</b> |

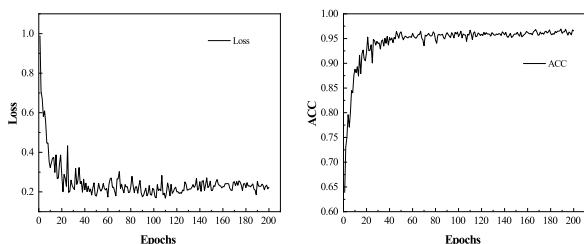
and trained on the underwater biological image dataset. The experimental results are shown in Table 5. SENet is a study of channel attention mechanisms for feature maps, and it is a study of attention mechanisms for convolutional kernels. Both are similar in that they enhance useful information and compress useless information.

In contrast, CBAM can generate attentional feature maps believed from both channels, and spatial dimensions have a stronger ability to improve the scalability of the underlying network. They can extract image feature information more effectively in complex underwater backgrounds. On the underwater biological image dataset, its accuracy reaches 96.93%. In conclusion, underwater images' channel and spatial attention combinations have higher classification accuracy.

**V. CONCLUSION**

In this paper, we propose a deep neural network model based on a dual attention mechanism for underwater biological image classification. Due to the complex underwater background and the situation that there are similar backgrounds in different categories, we choose to use the CBAM module to increase the saliency of image features. The combination and stacking of attention and convolution are realized to reduce feature parameters, thereby improving the learning performance and robustness of the model. Extensive experiments on underwater biological image datasets have shown that the DAMNet model obtains better results compared to other models. The model provides technical support in fishery and ocean exploration and deserves further study.

Although the classification results of the current models are good, these methods have certain limitations. First, the fish in the dataset is a large class, which has a poor learning effect and low classification accuracy. Second, there are limitations and singularity in the images for marine biometrics



**FIGURE 8.** Average loss curve. The trend loss of our model in training is consistent with the trend of training accuracy, which indicates that there is also no overfitting problem during the training period and the model has good generalization ability.

size of 16 and initial step size, average movement parameter, and learning rate set to 0.01, 0.9, and 0.01, respectively. The gravity optimizer speeds up the updating of weights by exponential translation, and the addition of the CBAM module enables the extraction of rich features in complex underwater backgrounds. The dual attention mechanism is employed to speed up the feature extraction and reduce the number of parameters to make the model converge faster and with lower losses. The average loss curve of the DAMNet model training is shown in Fig 8.

**E. ABLATION STUDY**

To further demonstrate the effective features obtained by combining dual attention mechanisms, this paper incorporates different attention mechanisms in the model for the ablation study. Attention mechanisms such as SENet [52], SKNet [53], and CBAM [54] were mainly compared, and different attention mechanisms were added to the model



identification, such as the presence of only one category on an image. In the model, as it is processed for underwater images, there is more feature information and noise information than land images, there are a large number of parameters, and the training time is longer. In our future work, we will add the fine classification of fish and will consider designing a model applied to underwater creature target detection to achieve the recognition of multiple targets in a single image.

## ACKNOWLEDGMENT

The authors would like to thank Xiwang Xie participated in the technical support of the paper.

## REFERENCES

- [1] M. Yang, J. Hu, C. Li, G. Rohde, Y. Du, and K. Hu, "An in-depth survey of underwater image enhancement and restoration," *IEEE Access*, vol. 7, pp. 123638–123657, 2019.
- [2] B. Taormina, M. P. Marzloff, N. Desroy, X. Caisey, O. Dugornay, E. M. Thiesse, A. Tancray, and A. Carlier, "Optimizing image-based protocol to monitor macroepibenthic communities colonizing artificial structures," *ICES J. Mar. Sci.*, vol. 77, no. 2, pp. 835–845, Mar. 2020.
- [3] Q. Li, X. Sun, J. Dong, S. Song, T. Zhang, D. Liu, H. Zhang, and S. Han, "Developing a microscopic image dataset in support of intelligent phytoplankton detection using deep learning," *ICES J. Mar. Sci.*, vol. 77, no. 4, pp. 1427–1439, Jul. 2020.
- [4] Y. Liu, Z. Yan, J. Tan, and Y. Li, "Multi-purpose oriented single nighttime image haze removal based on unified variational Retinex model," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Oct. 28, 2022, doi: 10.1109/TCSVT.2022.3214430.
- [5] W. Zhang, X. Pan, X. Xie, L. Li, Z. Wang, and C. Han, "Color correction and adaptive contrast enhancement for underwater image enhancement," *Comput. Electr. Eng.*, vol. 91, May 2021, Art. no. 106981.
- [6] W. Zhang, Y. Wang, and C. Li, "Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 718–735, Jul. 2022.
- [7] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-Laplacian reflectance priors," *IEEE Trans. Image Process.*, vol. 31, pp. 5442–5455, 2022.
- [8] X. Fu and X. Cao, "Underwater image enhancement with global-local networks and compressed-histogram equalization," *Signal Process., Image Commun.*, vol. 86, Aug. 2020, Art. no. 115892.
- [9] W. Zhang, L. Dong, X. Pan, P. Zhou, L. Qin, and W. Xu, "A survey of restoration and enhancement for underwater images," *IEEE Access*, vol. 7, pp. 182259–182279, 2019.
- [10] W. Zhang, L. Dong, and W. Xu, "Retinex-inspired color correction and detail preserved fusion for underwater image enhancement," *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106585.
- [11] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.
- [12] W. Zhang, L. Dong, T. Zhang, and W. Xu, "Enhancing underwater image via color correction and bi-interval contrast enhancement," *Signal Process., Image Commun.*, vol. 90, Jan. 2021, Art. no. 116030.
- [13] L. Bai, W. Zhang, X. Pan, and C. Zhao, "Underwater image enhancement based on global and local equalization of histogram and dual-image multi-scale fusion," *IEEE Access*, vol. 8, pp. 128973–128990, 2020.
- [14] X. Sun, J. Shi, J. Dong, and X. Wang, "Fish recognition from low-resolution underwater images," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, 2016, pp. 471–476.
- [15] N. M. S. Iswari, Wella, and Ranny, "Fish freshness classification method based on fish image using k-nearest neighbor," in *Proc. 4th Int. Conf. New Media Stud. (CONMEDIA)*, Nov. 2017, pp. 87–91.
- [16] B. V. Deep and R. Dash, "Underwater fish species recognition using deep learning techniques," in *Proc. 6th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Mar. 2019, pp. 665–669.
- [17] D. Zhang, D.-J. Lee, M. Zhang, B. J. Tippetts, and K. D. Lillywhite, "Object recognition algorithm for the automatic identification and removal of invasive fish," *Biosyst. Eng.*, vol. 145, pp. 65–75, May 2016.
- [18] K. Cheng, X. Cheng, Y. Wang, H. Bi, and M. C. Benfield, "Enhanced convolutional neural network for plankton identification and enumeration," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0219570.
- [19] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, Jan. 2021.
- [20] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnol. Oceanogr., Methods*, vol. 14, no. 9, pp. 570–585, Sep. 2016.
- [21] M. Khishe and M. R. Mosavi, "Classification of underwater acoustical dataset using neural network trained by chimp optimization algorithm," *Appl. Acoust.*, vol. 157, Jan. 2020, Art. no. 107005.
- [22] V. J. D. Almero, R. S. Concepcion, E. Sybingco, and E. P. Dadios, "An image classifier for underwater fish detection using classification tree-artificial neural network hybrid," in *Proc. Int. Conf. Comput. Commun. Technol. (RIVF)*, Oct. 2020, pp. 1–6.
- [23] A. Mahmood, M. Bennamoun, S. An, F. Soheli, and F. Boussaid, "ResFeats: Residual network based features for underwater image classification," *Image Vis. Comput.*, vol. 93, Jan. 2020, Art. no. 103811.
- [24] W. Zhang, L. Dong, X. Pan, J. Zhou, L. Qin, and W. Xu, "Single image defogging based on multi-channel convolutional MSRCR," *IEEE Access*, vol. 7, pp. 72492–72504, 2019.
- [25] Y. Xu, Y. Zhang, H. Wang, and X. Liu, "Underwater image classification using deep convolutional neural networks and data augmentation," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Oct. 2017, pp. 1–5.
- [26] Q. Qi, K. Li, H. Zheng, X. Gao, G. Hou, and K. Sun, "SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception," *IEEE Trans. Image Process.*, vol. 31, pp. 6816–6830, 2022.
- [27] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.
- [28] J. Xie, G. Hou, G. Wang, and Z. Pan, "A variational framework for underwater image dehazing and deblurring," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3514–3526, Jun. 2022.
- [29] C. Li, C. Guo, L. Han, J. Jiang, M.-M. Cheng, J. Gu, and C. C. Loy, "Low-light image and video enhancement using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9396–9416, Dec. 2022.
- [30] J. Jäger, M. Simon, J. Denzler, V. Wolff, K. Fricke-Neuderth, and C. Kruschel, "Croatian fish dataset: Fine-grained classification of fish species in their natural habitat," in *Proc. BMVC*, vol. 2, Swansea, Wales, 2015, pp. 1–7.
- [31] S. Villon, D. Mouillot, M. Chaumont, E. S. Darling, G. Subsol, T. Claverie, and S. Villéger, "A deep learning method for accurate and fast identification of coral reef fishes in underwater images," *Ecol. Informat.*, vol. 48, pp. 238–244, Oct. 2018.
- [32] A. A. Balakrishnan, M. S. Bijoy, and M. H. Supriya, "Classification of low quality underwater objects using convolutional neural networks and transfer learning," in *Proc. OCEANS*, Chennai, Feb. 2022, pp. 1–4.
- [33] M. Aridoss, C. Dhasarathan, A. Dumka, and J. Loganathan, "DUICM deep underwater image classification model using convolutional neural networks," *Int. J. Grid High Perform. Comput.*, vol. 12, no. 3, pp. 88–100, Jul. 2020.
- [34] X. Fu, Q. Qi, Z.-J. Zha, X. Ding, F. Wu, and J. Paisley, "Successive graph convolutional network for image de-raining," *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1691–1711, May 2021.
- [35] M. Yang, H. Wang, K. Hu, G. Yin, and Z. Wei, "IA-Net: An inception-attention-module-based network for classifying underwater images from others," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 704–717, Jul. 2022.
- [36] M. Paraschiv, R. Padrino, P. Casari, E. Bigal, A. Scheinin, D. Tchernov, and A. Fernández Anta, "Classification of underwater fish images and videos via very small convolutional neural networks," *J. Mar. Sci. Eng.*, vol. 10, no. 6, p. 736, 2022.
- [37] M. Mathur and N. Goel, "FishResNet: Automatic fish classification approach in underwater scenario," *Social Netw. Comput. Sci.*, vol. 2, no. 4, pp. 1–12, Jul. 2021.
- [38] Z. Jiang, C. Zhao, and H. Wang, "Classification of underwater target based on S-ResNet and modified DCGAN models," *Sensors*, vol. 22, no. 6, p. 2293, Mar. 2022.

[39] T. P. Marques and A. Branzan Albu, "L2UWE: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 538–539.

[40] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[41] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[42] M. Jahidul Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," 2020, *arXiv:2002.01155*.

[43] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[48] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[49] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying convolution and attention for all data sizes," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 3965–3977.

[50] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13733–13742.

[51] N. Park and S. Kim, "How do vision transformers work?" 2022, *arXiv:2202.06709*.

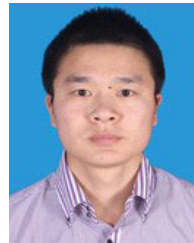
[52] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[53] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[54] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.



**LING ZHOU** received the M.S. degree from Huazhong Agricultural University, in 2019. She is currently a Teaching Assistant with the Henan Institute of Science and Technology. Her research interests include bioinformation technology and agricultural information.



**SONGLIN JIN** received the M.S. degree in computer application technology from Central China Normal University, in 2010. He is currently a Lecturer with the Henan Institute of Science and Technology. His research interests include machine learning and hyperspectral image processing.



**ZHENG LIANG** received the Ph.D. degree from the Information Science and Technology College, Dalian Maritime University, Dalian, China, in 2021. He is currently a Lecturer with Anhui University, Hefei, China. He has authored or coauthored more than 15 papers in some famous journals or conferences, including *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, *Neurocomputing*, and *KBS*. His research interests include computer vision and machine learning.



**WENYI ZHAO** is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include machine learning, deep learning, and machine vision.



**PEIXIN QU** received the M.S. degree in physical electronics from Belarusian State University, in 2007. He is currently an Associate Professor with the Henan Institute of Science and Technology. His research interests include the areas of machine learning and wireless sensor networks.



**TENGFELI** is currently pursuing the master's degree with the School of Information Engineering, Henan Institute of Science and Technology. His research interests include image processing and machine learning.



**WEIDONG ZHANG** (Member, IEEE) received the Ph.D. degree in information and communication engineering from the School of Information Science and Technology, Dalian Maritime University, Dalian, China, in June 2022. He is currently a Lecturer with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China. He has authored or coauthored more than 35 papers in some famous journals or conferences, including *IEEE TRANSACTIONS ON IMAGE PROCESSING*, *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *AAAI*, *IEEE JOURNAL OF OCEANIC ENGINEERING*, and *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*. His current research interests include image processing, computer vision, and machine learning.

...