WildFishNet: Open Set Wild Fish Recognition Deep Neural Network With Fusion Activation Pattern

Xiaoya Zhang, Baoxiang Huang, *Member, IEEE*, Ge Chen, Milena Radenkovic, *Member, IEEE*, and Guojia Hou

***Abstract*—Wild fish recognition is a fundamental problem of ocean ecology research and contributes to the understanding of biodiversity. Given the huge number of wild fish species and unrecognized category, the essence of the problem is an open set fine-grained recognition. Moreover, the unrestricted marine environment makes the problem even more challenging. Deep learning has been demonstrated as a powerful paradigm in image classification tasks. In this article, the wild fish recognition deep neural network (termed WildFishNet) is proposed. Specifically, an open set fine-grained recognition neural network with a fused activation pattern is constructed to implement wild fish recognition. First, three different reciprocal inverted residual structural modules are combined by neural structure search to obtain the best feature extraction performance for fine-grained recognition; next, a new fusion activation pattern of softmax and openmax functions is designed to improve the recognition ability of open set. Then, the experiments are implemented on the WildFish dataset that consists of 54 459 unconstrained images, which includes 685 known classes and 1 open set unrecognized category. Finally, the experimental results are analyzed comprehensively to demonstrate the effectiveness of the proposed method. The in-depth study also shows that artificial intelligence can empower marine ecosystem research.**

***Index Terms*—Deep neural network, fusion activation pattern, neural structure search, open set fine-grained recognition, wild fish recognition.**

I. Introduction

FISH are essential elements for the survival and development of other marine organisms and for promoting the circulation of marine substances [1] and play a critical role in maintaining the balance of marine ecosystems [2]. In addition, fish biodiversity is particularly important for the sustainable development of the ocean [3]. Nevertheless, due to human activities, such as overfishing and discarding marine debris, the survival of fish is seriously threatened [4], and global climate and ecological changes cause abnormal behavior of fish [5]. Consequently, achieving fine-grained identification of wild fish open set is significant for the efficient management of fishery resources, protection of marine biodiversity, and intelligent monitoring of marine environmental changes.

Traditionally, the fish image classification studies have been limited to closed-set recognition of known species [6], [7], whose training data and test data are derived from the same distribution. However, in the real open ocean environment, there are still a large number of unknown species of fish living in ocean areas that have never been explored by humans [8]. If the closed-set recognition algorithm is still used to classify the wild fish, the unknown class will be incorrectly divided directly into one of the known classes. Open set recognition (OSR) [9] is different from closed set recognition in that the test set contains categories not found in the training set [10]. In the open-set setting, not only the known classes of the test set must be correctly distinguished, but also the newly emerging classes should be detected [11]. Moreover, the multispecies classification of fish requires more refined identification due to the intraclass variation and interclass similarity of fish characteristics [12]. As a whole, the use of open-set fine-grained identification methods is more consistent with the actual marine environment. In the method based on machine learning, the 1-vs-Set Machine [9], which optimizes open space risk terms using the linear support vector machine (SVM) algorithm, Weibull-calibrated SVM (W-SVM) [13] based on the extreme value theory (EVT), and other traditional open set identification algorithms have been proposed.

Various emerging deep learning methods [14] have been applied to image recognition, among which convolutional neural networks (CNNs) are the mainstream [15] to tackle OSR [16] and fine-grained classification tasks [17], [18]. Bendale et al. [19] presented the Openmax algorithm instead of the softmax layer to calculate the pseudoactivation probabilities of unknown classes. Classification-reconstruction learning for OSR [20] reconstructed the latent feature representations extracted from the supervised classification process to better distinguish known and unknown classes. Neal et al. [21] applied

Manuscript received 30 May 2023; revised 13 July 2023; accepted 22 July 2023. Date of publication 28 July 2023; date of current version 10 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42276203 and Grant 42030406, in part by the Laoshan Laboratory under Grant LSKJ202204302, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2021MD001. *(Corresponding author: Baoxiang Huang.)*

Xiaoya Zhang and Guojia Hou are with the Department of Computer Science and Technology, Qingdao University, Qingdao 266071, China (e-mail: zxy17864220487@163.com; hgj2015@qdu.edu.cn).

Baoxiang Huang is with the Department of Computer Science and Technology, Qingdao University, Qingdao 266071, China, and also with the Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266228, China (e-mail: hbx3726@163.com).

Ge Chen is with the Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266228, China, and also with the School of Marine Technology, Institute for Advanced Ocean Study, Ocean University of China, Qingdao 266075, China (e-mail: gechen@ouc.edu.cn).

Milena Radenkovic is with the School of Computer Science and Information Technology, The University of Nottingham, NG8 1BB Nottingham, U.K. (e-mail: milena.radenkovic@nottingham.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2023.3299703

generative adversarial networks to synthesize fake unknown class training samples for OSR. In addition, the fine-grained recognition methods based on the deep learning framework mainly include region localization techniques [22], [23], network integration [24], [25], higher order feature coding [26], [27], and so on. Artificial intelligence has gradually been more applied in the field of marine science [28], [29], [30]. Roberts et al. [31] introduced SVMs as the underlying classifier to establish the classification of marine fish. Chen et al. [32] used the context information from the scenario around to obtain improved fish classification effect. Bilinear CNNs [33] based on part localization are applied to fish classification to promote fine-grained feature extraction capability. Li et al. [12] designed a multiscale fusion feature model to identify 15 types of highly similar fish and achieved an accuracy rate of 95.31%. The aforementioned studies can demonstrate that the simple CNN has deeper and more accurate feature expression capabilities than the traditional methods, and also is more suitable for identifying wild fish images. For the OSR of wild fish, Zhuang et al. [34] was the first to use the residual network (ResNet) introduced by the Openmax algorithm to achieve open-set fine-grained identification of wild fish. Akhtarshenas and Toosi [35] improved the accuracy of fish in open scenes by using an autoencoder to prevent unknown class samples with large reconstruction errors from entering the classifier. However, there will be some image quality defects when expanding the samples of known classes to generate wild fish instances of unknown classes. Due to the complexity of wild fish features, the reconstructed features are relatively unstable for the discrimination of unknown wild fish. Based on the aforementioned analysis, it is still a challenge for wild fish in realizing open-set fine-grained recognition.

Benefiting from the accumulated wild fish benchmark datasets [34], these existing methods have achieved efficient recognition results, but there are still some issues worth considering, in other words, the main contributions of this article are as follows.

1) The complicated marine environment and feature diversity yield the high difficulty of wild fish recognition, and the unknown category is also a challenge. Consequently, a deep neural network called WildFishNet is constructed to perform open-set fine-grained recognition of wild fish.
2) Inspired by EfficientNetV2 [36] architecture, the distinguishable features can be extracted with different reciprocal inverted residual structural modules to enhance the fine-grained classification efficiency.
3) The closed set softmax activation mechanism and open set Openmax activation algorithm are combined to improve the recognition accuracy.
4) Three common classification networks, including ResNet50 [37], MobileNetV2 [38], and EfficientNetV2, are employed as the network backbone for comparison, and the proposed fusion activation classification mechanism is compared with other typical classifiers experimentally. The experimental results indicate that the method has strong data fitting ability and better classification effect than other state-of-the-art methods.

The rest of this article is organized as follows. Section II illustrates the dataset and problem formulation. The proposed method is introduced in Section III, including the neural network architecture and the open set fine-grained classification mechanism. Section IV clarifies the experimental results. Finally, Section V concludes this article.

II. Preliminaries

Open set fine-grained wild fish recognition mainly includes two tasks: classification and identification. The classification process is to classify the known types of wild fish into multiple categories according to their characteristics. While the identification problem is to detect samples different from the training set distribution categories from the test set, which is essentially a binary classification problem of known classes and unknown categories. Consequently, the recognition of the wild fish open-set is an image multiclassification problem.

A. Dataset

Global wild fish resources are abundant, but the number of images or fish categories included in current fish datasets are relatively small, such as Fish4Knowledge [39], QUT dataset [40], Labeled Fish in Wild [41], etc. The relevant data in this study is the WildFish dataset [34].

The WildFish dataset consists of 54 459 fish images, containing 1000 species of wild fish, collected from a variety of specialist websites and a Google engine based on FishBase, a global fish biodiversity information system. After removing uncertain and overlapping images from different sources, each category ended up with at least 30 fish images. The construction of WildFish dataset can maintain the authenticity and reliability of wild fish images. The number of species has a greater impact on the classification performance, so WildFish dataset is more conducive to the generalization ability of experimental models than other fish datasets.

In this case, this study utilizes 685 species of wild fish in the dataset as known classes used to train the feature extractor of the model. Meanwhile, the remaining wild fish classes are retained as unknown classes in the test open-set to evaluate the OSR method. Intraclass variation and interclass similarity in wild fish resulted in a high diversity of traits, with some known and unknown classes of wild fish being extremely similar, as shown in Fig. 1. This fact poses research challenges for OSR of wild fish. The aforementioned points indicate that the WildFish dataset is more consistent with the complexity of the real scenes.

B. Problem Formulations

To better describe the mathematical process, we formalize the problem of OSR in wild fish. First, let $\boldsymbol{\chi} = D_s \cup D_t$ denotes the open set data, where the number of categories of the source domain sample set D_s and the target domain sample set D_t is not consistent. $D_s = \{(x_i, y_i)\}_{i=1}^{n}$ represents n annotated input images, where $y_i \in C$ and C is the set of categories of known samples. $k = \{1, 2, \ldots, K\}$ is defined as a vector of K class labels, and here, $K = 685$ denotes 685 known species of wild

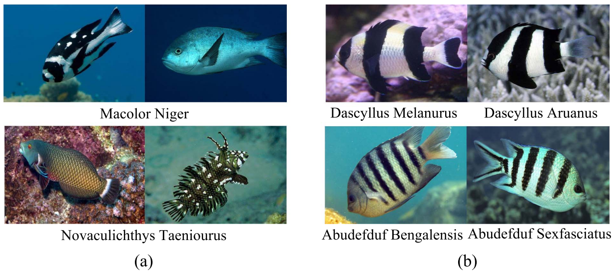


Fig. 1. Feature diversity of the Wildfish dataset. Each group in (a) and (b) demonstrates intraclass variability and interclass similarity of wild fish, where Abudefduf Bengalensis and Abudefduf Sexfasciatus belong to unknown and known classes, respectively. (a) Images of intraclass fish. (b) Image of interclass fish.

fish. Note that in OSR, there is one more unknown category $\{U\}$ in the target domain, so $y_j \in (C \cup \{U\})$ in the test set $D_t = \{x_j, y_j\}_{j=1}^{m}$. The purpose of recognition is to assign probability scores $s(y_{(k)} \in C|x_k)$ of each known class to each fish image $x_k \in \chi$, so as to instruct the classifier to correctly predict the class of the sample and judge whether the sample belongs to the unknown class.

In the closed set multiclassification task, to increase the difference between probabilities and ensure the exclusivity of classification, softmax activation function is used to normalize the probabilities of classification labels, and the sum of the output probabilities of all classification labels is 1. The softmax function is described as

$$P\left(y_{(i)}|i\right) = \frac{\exp(s\left(y_{(i)}|i\right))}{\sum_{k=1}^{|C|}\exp(s(k|i))} \tag{1}$$

where $P(y_{(i)}|i)$ denotes the probability that the ith training sample is predicted to be the $y_{(i)}$th class; $s(y_{(i)}|i)$ is the output fraction of the $y_{(i)}$th class to which the ith depth feature belongs, and $|C|$ is the number of classes of the known class.

Generally speaking, the cross-entropy loss function is used as the objective function of the network. However, only the separability between classes is realized, and the intraclass distance is not optimized. So we introduce center loss [42] into the OSR of wild fish to reduce the intraclass feature spacing. The known classification of fish occupies only a portion of the total feature space and leaves enough area for new class centers. This network combines cross-entropy loss function and central loss function to reflect the accuracy of probabilistic classification, which can be abstractly formulated as

$$\text{Loss} = -\sum_{i=1}^{Nb}\log(P\left(y_{(i)}|i\right)) + \frac{\lambda}{2}\sum_{i=1}^{Nb}\left\|x_i - c_{y(i)}\right\|_2^2 \tag{2}$$

where $x_i \in \Re^d$ represents the feature before the fully connected layer, d is the feature dimension; and $c_{y(i)}$ means the feature center of the $y_{(i)}$th category. The Euclidean distance from each training sample to the center of its category is calculated. The smaller the sum of distances, the more compact the intraclass feature. Nb is batch size; λ represents the hyperparameter used to balance the two losses. To verify the usefulness of center loss in OSR, the feature distributions of open-set samples are visualized in Section IV-B according to different settings of the hyperparameter λ.

III. Methodology

A. Wild Fish Recognition Neural Network

The overall structure of the proposed wild fish open-set recognition framework WildFishNet is shown in Fig. 2. It mainly consists of two parts: feature extractor and classifier. A series of low-dimensional and high-dimensional feature vectors are extracted from the wild fish images to be classified through model convolution training, and then, predicted by the classifier. For OSR, there is a positive correlation between the accuracy of the classifier in the closed-set class and the performance in the open set [43]. Therefore, two activation mechanisms, softmax and openmax, are fused as the classifier of the test model to achieve the correct classification of known wild fish and the detection of unknown fish in this study.

B. Discriminative Feature Extractor

Based on EfficientNetV2, part of the structure is modified to fit the extraction of complex features from multiple classes of wild fish. As the network backbone, the feature extractor is mainly composed of a series of three different inverted residual modules Fused-MBconv block, SFused-MBConv block, and MBConv block, whose detailed structure includes the convolutional layer, attention module, and dropout layer, as illustrated in Fig. 2. The input size of the network follows the default size of EfficientNetV2. First, the training image (image size: $128 \times 128 \sim 300 \times 300$) is preliminarily processed by 3×3 ordinary convolution, and then, features are extracted by stacking three groups of inverted residual modules for different times. Different from the residual structure, each block in the inverted residual structure uses different multiplier factors to expand its width to extract more feature information, and finally, reduces the number of channels. Batch normalization (BN) [44] and sigmoid linear unit (SiLU) activation functions [45] are performed after convolution. Similarly, the test image of size $384{\times}384$ flows through the network backbone to obtain corresponding features.

The detailed construction of the inverted residual module is shown in Fig. 3. It can be observed from Fig. 3(b) that SFused-MBConv block is embedded with the spatial attention module [46] based on the Fused-MBconv block. It compresses the features of the whole spatial domain through global pooling and average pooling and then uses $3{\times}3$ convolution to extract large-grained information from the feature map after concatenation. Finally, the activation function maps the 2-D spatial attention map into the weight probability of [0,1] to reflect the different importance of image information. In Fig. 3(c), the MBConv block uses the squeeze-and-excitation (SE) module to learn the nonlinear relationship between each channel after obtaining features by depthwise convolution, which is conducive to the subsequent classification and recognition of wild fish. Furthermore, the shortcut connections [37] in the inverted

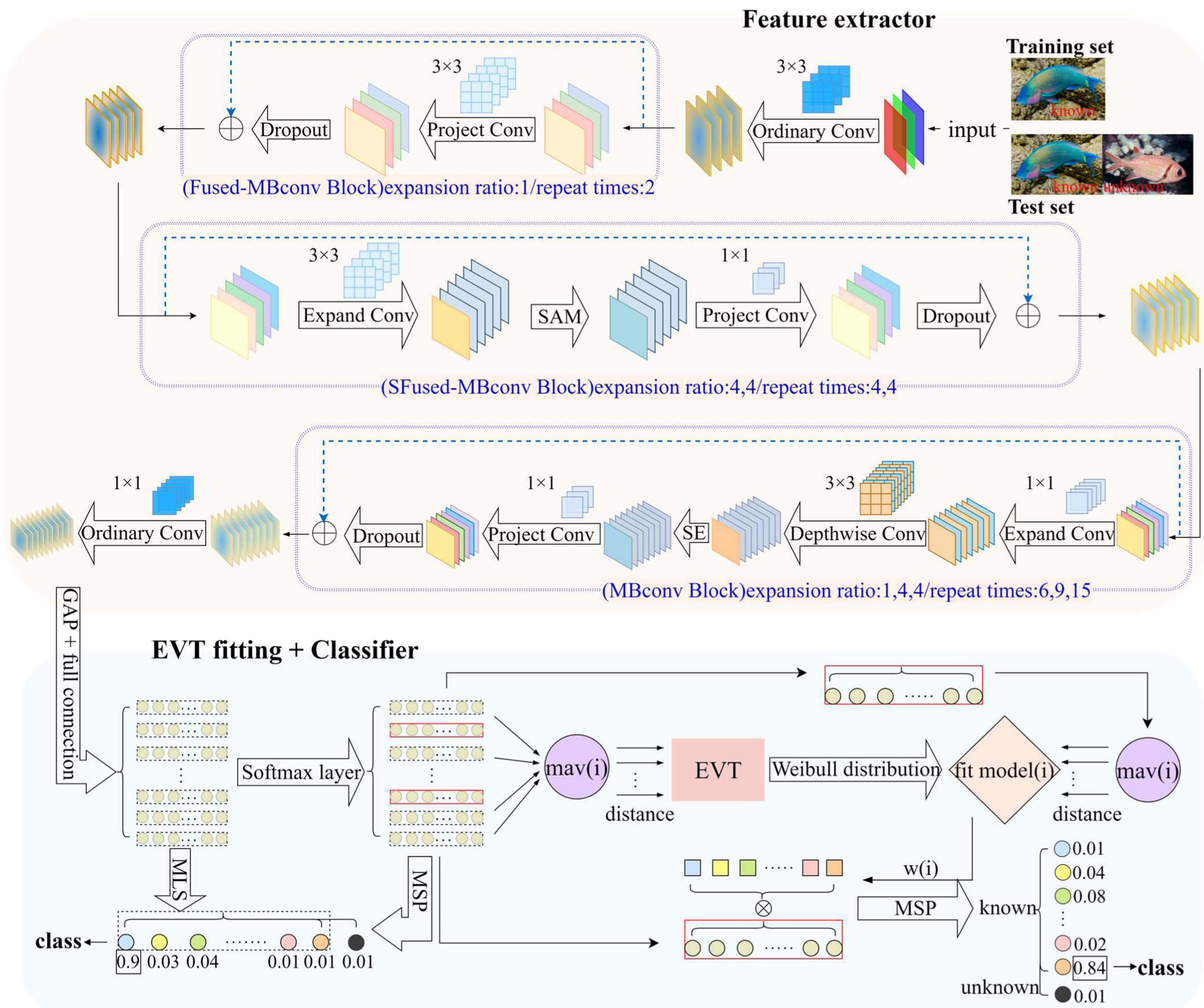


Fig. 2. Overall framework of the WildFishNet structure.

residual block add up the feature maps obtained in different ways so that the low-level information will not disappear with the deepening of the network structure. Simultaneously, the dropout layers [47] adopt the survival probability principle of linear decay to randomly drop network blocks to prevent overfitting. The combination of shortcut connections and dropout layers optimizes the feature extraction process of deep networks. Therefore, WildFishNet can extract rich feature information from wild fish images with complex features, and also allows to finally have a lower prediction time.

C. Open Set Classifier With a Fusion Activation Pattern

As shown in Fig. 2, the linear classifier (full connection layer) of the network backbone is used as the classification layer during the training stage. The learned feature vectors go through the classification layer to obtain the activation vectors containing 685 logits, and we directly adopt the maximum logit score baseline as the criterion for judging the categories. For the test classification process, the main idea is to use softmax classifier to perform preliminary screening after the feature extraction of the input test image, and then, introduce openmax layer to correct the activation vector of the misclassified sample, and finally, distinguish the category of the image.

1) Softmax Layer: Each test data passes through the feature extractor to get a corresponding activation vector, which contains the logit belonging to each known category. The activation vector of the ith data is denoted as $\mathrm{av}_i = \{\mathrm{logit}_1, \mathrm{logit}_2, \ldots, \mathrm{logit}_c\}$, where c is the number of known classes. The activation vector of data are converted into probability vector $\mathrm{pro}_i = \{p_1, p_2, \ldots, p_c\}$ by softmax function. Finally, according to the maximum softmax probability (MSP) baseline and the set probability threshold δ, the preliminary classification is carried out. The classification principle of the softmax layer is defined as

$$g_i = \begin{cases} \mathrm{argmax}(\mathrm{pro}_i), & \max(\mathrm{pro}_i) \geq \delta \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where g_i represents the initial classification result of the ith data. When the highest probability is less than the threshold δ, the

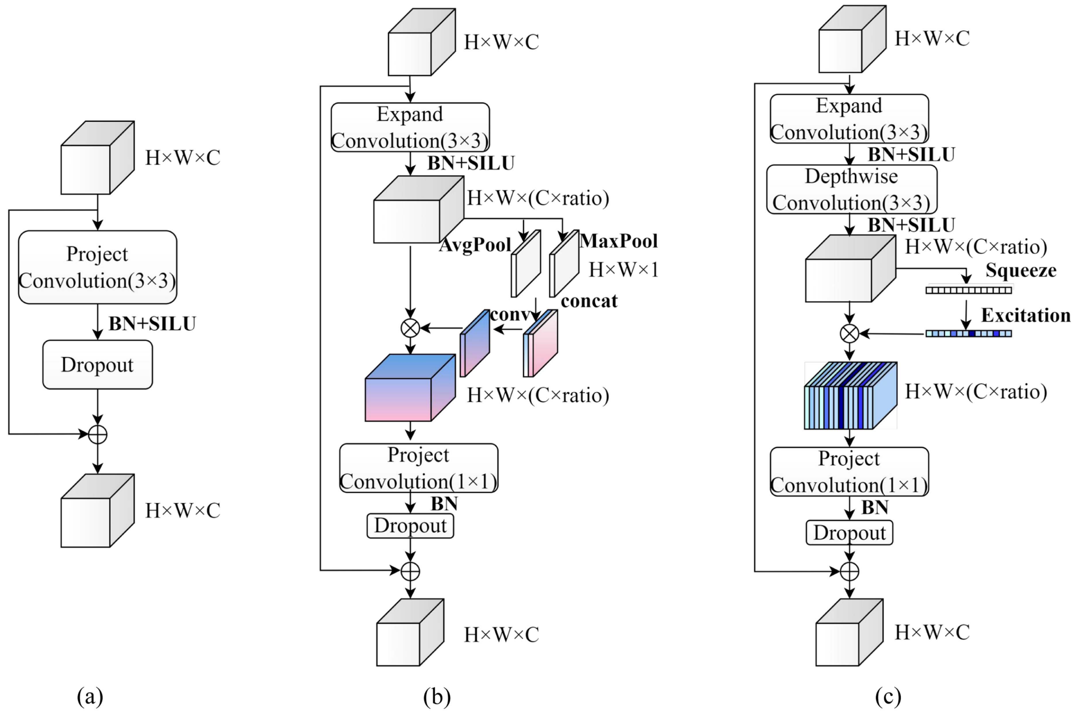


Fig. 3. Detailed flowchart of the inverted residual structure module. (a)–(c) Main modules of the feature extractor in WildFishNet. According to the NAS technique, each module has a different expansion ratio factor in the channel dimension and the number of stacking times in the depth dimension. (a) Fused-MBconv block. (b) SFused-MBconv block. (c) MBconv block.

image is directly divided into the unknown class; otherwise, the prediction result is the class to which the highest classification probability belongs. Finally, the correct classified data are retained, and the rest of the wrong ones are input to the openmax layer.

2) EVT Distribution Fitting: To solve the open set problem, the openmax algorithm utilizes the EVT to model the distance of these activation vectors from the mean of each class to generate updated activation vectors.

First, the activation vectors of training data with the correct classification prediction are calculated and retained, and the vectors are divided according to the data label. Here, the vector set of the kth known class data is denoted as $\mathrm{AV}_k = \{\mathrm{av}_1, \mathrm{av}_2, \ldots, \mathrm{av}_m\}$, where m means that m data in the kth class training data are identified as the kth class. The mean of the corresponding vectors of each category is used as the category center of the category, denoted as MAV. The MAV of class k is described as

$$\mathrm{MAV}_k = \frac{1}{m}\sum_{t=1}^{m} \mathrm{av}_t. \tag{4}$$

Then, the Euclidean distance between the activation vector and its category center is calculated for each data. It is noted that the loss is constructed based on the Euclidean distance for optimization in the training phase, so the measurement method in the test is consistent with that used. Based on EVT, the distances of all kinds of samples are sorted, and the tail maximum distances are analyzed to satisfy Weibull distribution [19]. Finally, the fitted distribution model for each known class is obtained.

3) Openmax Layer: The openmax layer is introduced to correct the activation vector before softmax layer of the test data to realize the final open-set recognition task. If the logit of category k to which the jth error data belongs is modified, the distance from the activation vector of the data to the MAV of the kth category is first calculated. Then, the distance is input into the the cumulative distribution function (CDF) of the Weibull distribution [12] to calculate the probability correction weight w_k of class k. The CDF is described as

$$w_k = 1 - \frac{\alpha - k}{\alpha}\exp\left\{-\left(\frac{\mathrm{av}_j - \gamma_k}{\eta_k}\right)^{\beta_k}\right\} \tag{5}$$

where α is the scale parameter, and $\{\gamma_k \in R^c, \beta_k \in R, \eta_k \in R\}$ are the parameters obtained by fitting the Weibull distribution. The logits in the activation vector are updated by using w_k. The calculation of logit'_k belonging to the known class k and logit_u of the unknown class (the probability that an image does not belong to any known class) in the vector are shown in the following equations, respectively:

$$\mathrm{logit}'_k = \mathrm{logit}_k \cdot w_k \tag{6}$$

$$\mathrm{logit}_u = \sum_{k=1}^{c} \mathrm{logit}_k \cdot (1 - w_k). \tag{7}$$

The final corrected activation vector $av'_j = \{\mathrm{logit}'_1, \mathrm{logit}'_2, \ldots, \mathrm{logit}'_c, \mathrm{logit}_u\}$ is normalized to obtain the new probability

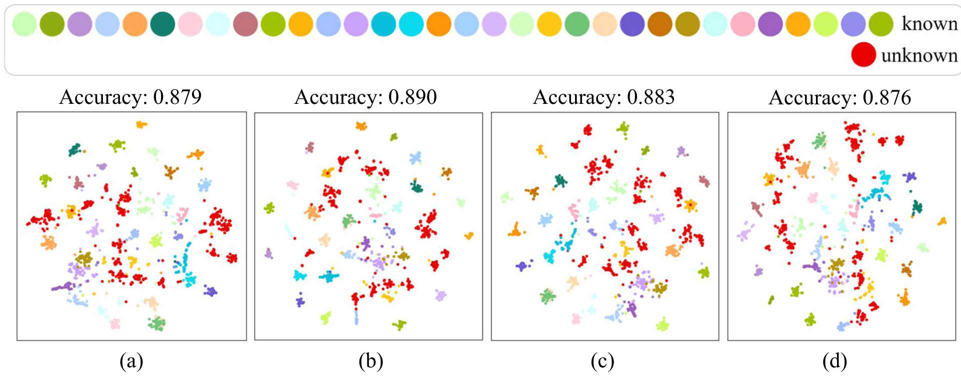


Fig. 4. (a)–(d) 2-D feature distribution of the t-SNE visualization generated by WildFishNet under the condition that the center loss hyperparameter λ takes different values. (a) denotes the semantic features of the data with only cross-entropy loss in the model objective function. (a) $\lambda = 0$. (b) $\lambda = 0.001$. (c) $\lambda = 0.0001$. (d) $\lambda = 0.00001$.

vector $\text{pro}'_j = \{p'_1, p'_2, \ldots, p'_c, p_{c+1}\}$. Likewise, the distribution threshold ρ is set and combined with the MSP baseline to achieve the reclassification of faulty data. If the maximum classification probability is obtained in the unknown class or the maximum classification probability is less than threshold ρ, the jth error data are identified as the wild fish of the unknown category. On the contrary, the jth error data are classified as the known wild fish with the highest classification probability.

IV. Experimental Results and Analysis

A. Experiments Implementations

The proposed architecture is implemented by PyTorch, and runs on a computer equipped with a GeForce RTX 2080Ti GPU with a capacity of 11 G. 60% of the known class wild fish images in the WildFish dataset are used for training and validation, and the remaining 40% and all unknown class data are used for testing. The network adopts a progressive learning method, which adaptively adjusts the image size and data enhancement to accelerate the learning of feature representation and prevent overfitting. The maximum input image size is set to 300×300. In the closed-set training process, WildFishNet adopts transfer learning and loads the weights pretrained in ImageNet to accelerate the convergence. The batch size of 16 is used to train the parameters of all layers in the network. Stochastic gradient descent is selected as the model optimizer. The initial learning rates of cross entropy loss and center loss are 0.01 and 0.1, respectively. And the learning rate of cross entropy loss function is attenuated to the minimum value by the cosine annealing algorithm. Therefore, the learning rate decreases slowly in the initial and final stages of training and decreases rapidly in the middle stages. This strategy makes the model converge faster and achieves the best results.

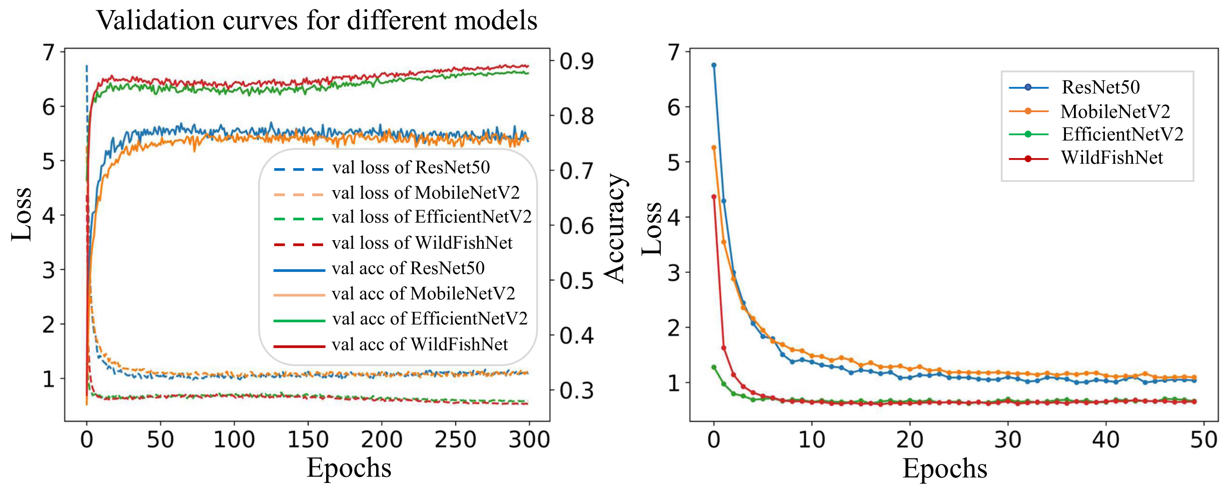
B. Parameters Sensitivity Analysis

To intuitively demonstrate the impact of the introduction of center loss and the different settings of hyperparameter λ on the OSR of wild fish, we extract 33 types of test data (including unknown class) according to the ratio of known and unknown classes in the test set. The feature distribution of test data in the last layer of WildFishNet was visualized, as shown in Fig. 4. Red dots and other colored dots represent unknown and known classes in the target domain, respectively. The t-SNE dimensional reduction visualization results of the feature representation show that the samples of the same class are more aggregated after adding the center loss, while different classes are discriminative. Looking closely at Fig. 4, as the value of the hyperparameter λ is set smaller, the features of the known classes become more divergent. As given in Fig. 4(b), the intraclass feature distance of the known class data is minimized when $\lambda = 0.001$, so there is enough feature space for the unknown class. It is obvious that there is an advantage at the stage of larger λ values, and the feature space of the sample reaches the best state at $\lambda = 0.001$. However, the features of individual unknown class data are wrongly close to those of known classes, which may be due to the similarity of features between known and unknown class data.

Furthermore, Fig. 4 also reports the accuracy of the validation data at different λ values, which is consistent with the feature changes caused by different λ settings. So, the hyperparameter λ of the center loss recommended in this article is 0.001, and this setting is used in the following experiments. Such comparisons demonstrate the effectiveness of the center loss, which not only benefits the recognition of unknown classes but also improves the classification accuracy of known classes.

C. Recognition Results

To explore the effect of different backbone networks on open-set recognition accuracy, we unbiasedly train ResNet50, MobileNetV2, EfficientNetV2, and our WildFishNet on maximum logit score classification baselines. Resnet50 is a CNN composed of a series of residual modules and solves the degradation problem in the deep network. MobileNetV2 is a lightweighted network that uses a depthwise separated convolutional structure. It is an easy-to-implement network with a simple structure and is also used as a basic layer in areas, such as detection and segmentation. The variation of accuracy and loss during training for each model is depicted in Fig. 5. The loss curve in Fig. 5(b)



(a) (b)

Fig. 5. Validation accuracy and loss curves for each model. (a) Graph of the whole validation process. (b) Loss curve of the first 50 epochs.

shows that EfficientNetV2 and WildFishNet converge faster than other models, and it can be seen from Fig. 5(a) that the difference between the predicted value of WildFishNet and the real value is the smallest.

At the same time, we deliberately select images of wild fish in different scenarios and input them into the trained models for testing to intuitively show the differences in the feature extraction of target wild fish by each model. As shown in Fig. 6, the class activation heatmap of the corresponding layer before the linear classifier of each model is visualized by Grad-CAM. For a given category of images, the heat map can reflect the regional features that the network focuses on, and the color of the region's marking represents the strength of the correlation. Due to the particularity of the open environment, the protective color or pattern of wild fish is usually extremely similar to the surrounding living environment, which makes the model overconsidered the background factor and ignores the main body of the fish. Likewise, the interclass similarity between known and unknown classes also causes the model to fail to determine their difference. It can be seen from Fig. 6 that the heat regions of ResNet50 and MobileNetV2 cannot accurately locate or completely cover the target wild fish, but the heat maps of WildFishNet and EfficientNetV2 are the most correct areas of concern when identifying fish and fish schools. Although WildFishNet is worse than EfficientNetV2 in extracting features from a single fish, it has a stronger recognition of target fish in groups of the same or different species. Thus, the features extracted by the last module of the three inverted residual structures in WildFishNet are shown in Fig. 7, and it can be found that the attention added in the SFused-MBConv module is more conducive to the extraction of spatial information for better localization of fish in complex images than the other modules. It can be said that WildFishNet is more consistent with the living habits of wild fish in complex marine environments.

In addition, multiclassification evaluation indicators are used, including the Precision as formula (8), the Recall as formula (9), and the F-macro value as formula (10) to comprehensively evaluate the performance of different models in the OSR task as follows:

$$\text{Precision}_i = \frac{\text{True positives}}{\text{predicted as positives}} \tag{8}$$

$$\text{Recall}_i = \frac{\text{True positives}}{\text{actual positives}} \tag{9}$$

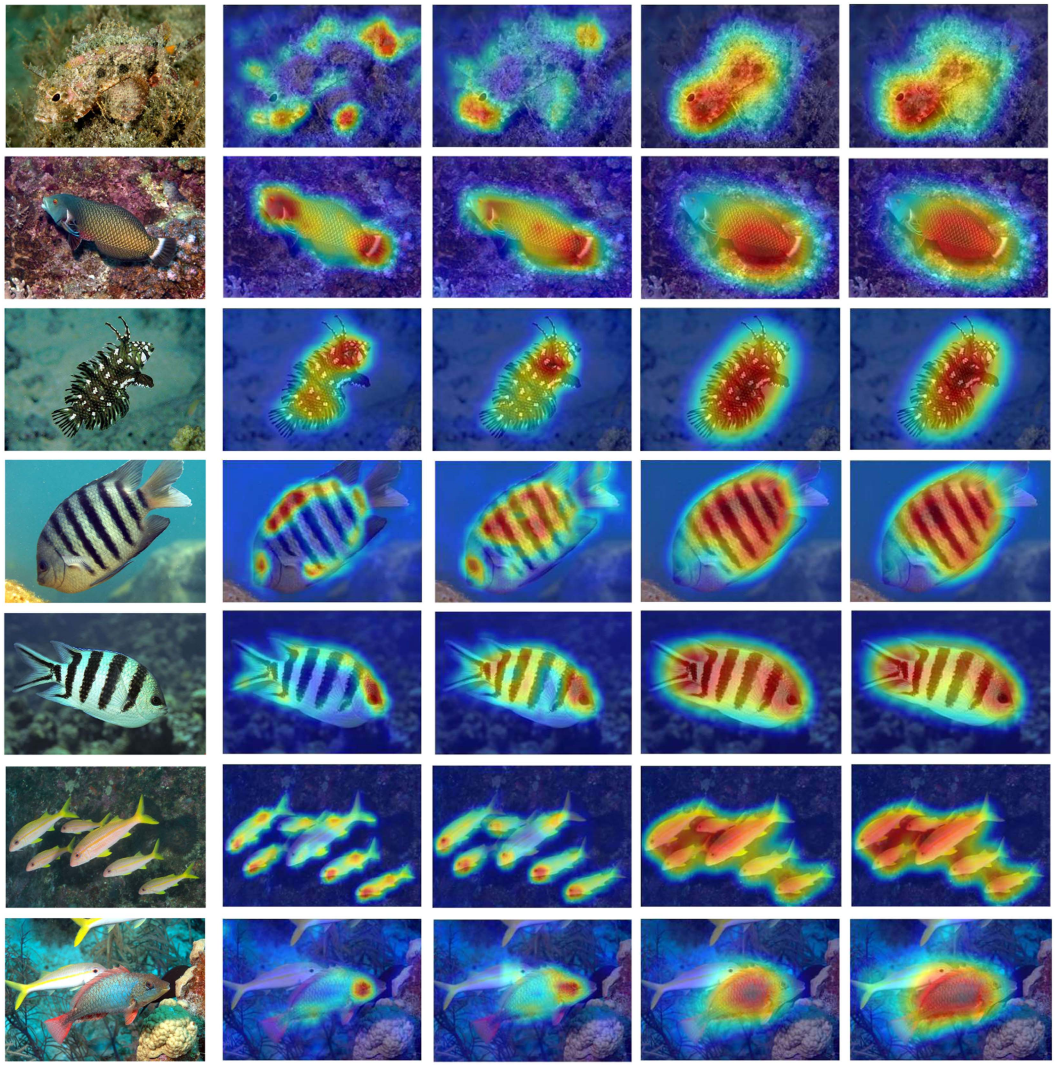
$$F\text{-macro} = \frac{1}{n}\sum_{i=1}^{n} F1_i = \frac{1}{n}\sum_{i=1}^{n} \frac{2 \bullet \text{Precision}_i \bullet \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{10}$$

where $i = 0, 1, 2, \ldots, O$, O is the category label of test data in the wild fish open set, specifically, $O = 685$. Precision$_i$ and Recall$_i$ represent the proportion of correctly classified wild fish in the predicted ith class and the true ith class, respectively. $F1_i$ represents $F1$-score of the ith category and the F-macro is the average of the sum of $F1$-*score* for each class.

Table I visually depicts the evaluation indicators and time that the test set is predicted by different models following the MSP baseline criteria. Although WildFishNet has more prediction time than ResNet50 and MobileNetV2, it greatly improves accuracy and F-macro, which means it identifies each species of wild fish most correctly among all models. At the same time, due to the improvement of the inverted residual modules,

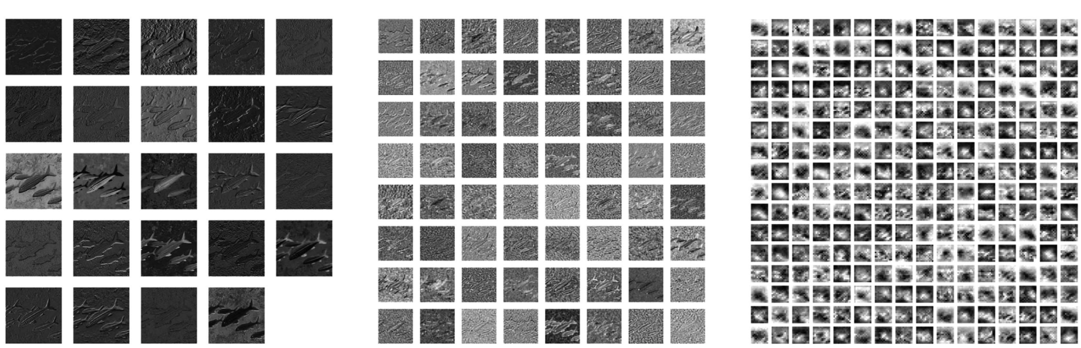
TABLE I
COMPARISON OF EVALUATION INDICATORS AND PREDICTION TIME FOR DIFFERENT MODELS

Indicators / Models	Accuracy	F-macro	Time(s)
ResNet50	67.8	67.3	525.0
MobileNetV2	67.9	66.1	509.5
EfficientNetV2	78.9	80.5	705.2
WildFishNet	80.1	81.5	680.5



(a) (b) (c) (d) (e)

Fig. 6. Grad-CAM heat maps of different models for a given wild fish. The heat map can reflect the focus area of the original image after feature extraction by the network, so we performed Grad-CAM for the output of the last convolution layer of each model. The figures indicate a total of seven types of wild fish images (a) and heat maps (b)–(e) of four networks for each type of image. (a) Original image. (b) ResNet50. (c) MobileNetV2. (d) EfficientNetV2. (e) WildFishNet.



(a) (b) (c)

Fig. 7. Feature visualization of different inverted residual modules in WildFishNet. (a)–(c) represent the feature maps of all channels of a layer in the three inverted residual modules. (a) Fused-MBconv block. (b) SFused-MBconv block. (c) MBconv block.

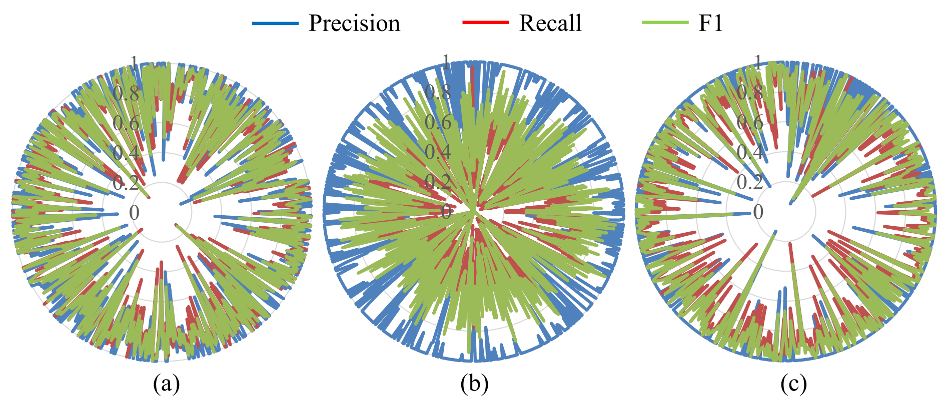


Fig. 8. Performance of different indicators for each category of the target data tested under the three classification methods. (a) Softmax. (b) Openmax. (c) Fused activation pattern.

TABLE II
EVALUATION INDEXES OF THE THREE CLASSIFICATION METHODS

Methods \ Indicators	Accuracy(known)	F1(unknown)
Softmax+Openmax	84.6	87.8
Softmax	82.1	76.2
Openmax	46.1	69.0

WildFishNet promotes classification efficiency compared with EfficientNetV2.

D. Different Activation Mechanisms Comparison

Softmax classifier has the advantages of fast speed and simple algorithm in open set multiclassification tasks. Usually, the objective function of multiple classification problems is cross entropy loss. Although the softmax classifier can only output the highest confidence in the category within the training distribution to match unknown data, the softmax function makes the influence of features on the probability multiplicative, so the highest confidence of unknown data belonging to the known class is relatively small. Consequently, softmax classifier can also identify some unknown data correctly by using a probability threshold. Furthermore, The classic openmax layer is also used as a comparison classifier. Openmax directly calculates the probability value of the unknown class to which the data belongs for the first time by refitting the distribution with EVT. Consequently, our proposed fusion activation classification method combines the aforementioned two classifiers to realize the open-set classification research of wild fish. Here, the ablation experiments are conducted on WildFishNet to observe the role of various parts of the fusion activation mechanism.

Table II records the known class accuracy and unknown class F1 values associated with the open set performance of WildFishNet's feature extractor combined with three classification methods on the test set. It can be observed that our method has a significant improvement in the recognition of unknown classes compared to keeping only the softmax classifier. And compared with the way of only using the openmax algorithm, it is confirmed that the contribution of the softmax layer in the closed set classification is larger. The experimental results show that the fusion activation classification mechanism not only maintains the performance in the classification of known classes but also achieves promising results in the recognition of unknown classes. Furthermore, Fig. 8 describes the performance of the target data identified by the three classification methods. Carefully observing Fig. 8(b) and (c), the recall curve for the test data identified using the openmax classifier converges more toward the center of the circle than those identified using the fusion activation pattern, indicating that the recall for most species using the openmax classification is low. Similarly, the precision of wild fish images for most species using softmax classification is not as high as that obtained by the fusion-activated classification. The F1 values take into account both the recall and precision, which can most intuitively show the recognition ability of each classification method. Obviously, the proposed fused activation pattern is more compatible with open-set fine-grained recognition of multispecies wild fish than other classifiers.

Finally, we observed the performance evaluation of each overall framework in wild fish open set recognition by changing the values of probability threshold δ and distribution threshold ρ in the fusion activation classification reference, as shown in Fig. 9. Compared with other architectures in Fig. 9(a) and (b), our architecture maintains the highest level in both the fine-grained classification of known classes and the recognition of unknown classes; in Fig. 9(c) and (d), when the probability threshold δ and the distribution threshold ρ are 0.75 and 0.4, respectively, the accuracy and F-macro of the open set reach the best values. Meanwhile, each classification metric of WildFishNet is the most robust under the condition of two threshold changes in the classifier. In addition, we compare the proposed method with other state-of-the-art methods on the WildFish dataset, as shown in Table III. Although the increase in the number of training species improves the complexity of the features, WildFishNet is 21.5% and 3.3% higher than ResNet50-OM and AE-CNN for the same number of species in the training set, respectively. Therefore, WildFishNet is the most promising

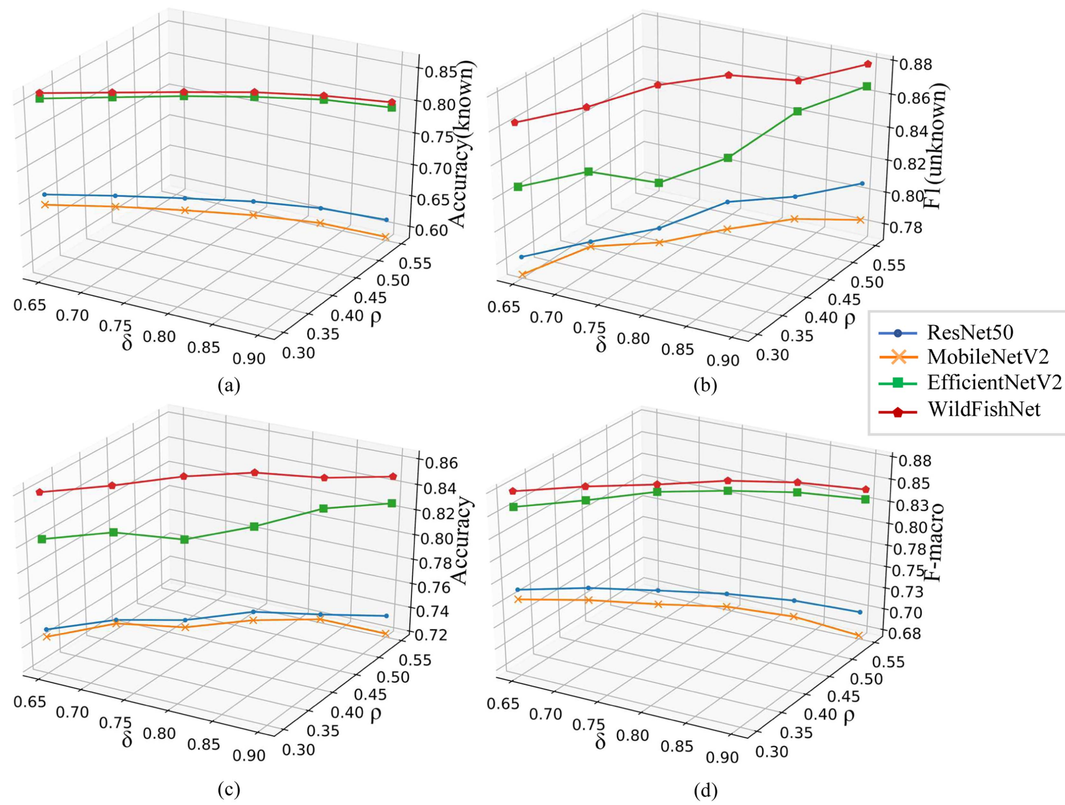


Fig. 9. Test performance of different models under two threshold changes for fused activation classification. (a)-(d) are Accuracy(known), F1(unknown), Accuracy and F-macro respectively.

TABLE III
PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH STATE-OF-THE-ART METHODS

Methods	ResNet50-OM [35]	AE-CNN [36]	WildFishNet	
	685 species	700 species	685 species	700 species
Accuracy	64.8	81.6	86.3	84.9

solution for open-set fine-grained classification of wild fish.

V. Conclusion

Wildfish recognition is typically the open set fine-grained classification problem. In this article, the proposed WildFishNet can achieve satisfying recognition results. The main idea can be expressed as follows. First, the feature extraction ability of the network is improved by adding an attention mechanism in the backbone; second, a hybrid loss function is adopted to reduce the intraclass feature distance to improve performance; then, a new fusion activation pattern of softmax and openmax functions is designed specifically; and finally, the recognition effectiveness of the proposed method is demonstrated by comprehensive experiments.

Moreover, we conduct ablation experiments on the proposed fusion activation classification method. The superiority of the classification benchmark has been verified in the experimental results, and it is more suitable for open-set classification tasks than other classifiers. Compared with the other four network backbones, our network achieves the highest accuracy of 86.3% in wild fish open-set multiclassification, the classification results are not perfect due to the high similarity of features between various wild fish, especially for known and unknown categories. The potential future research could extend to an in-depth analysis of wild fish in relation to marine ecological factors such as chlorophyll and microplastics.

Acknowledgment

In the spirit of reproducibility, the dataset and Python code of the proposed recognition methodology can be accessed with the linkage WildFishNet.

References

[1] X. Sun et al., “Transferring deep knowledge for object recognition in low-quality underwater videos,” *Neurocomputing*, vol. 275, pp. 897–908, 2018.
[2] B. W. Bowen, B. B. Collette, D. E. Facey, and G. S. Helfman, *The Diversity of Fishes: Biology, Evolution and Ecology*. Hoboken, NJ, USA: Wiley, 2022.
[3] H. Mehdi et al., “Municipal wastewater as an ecological trap: Effects on fish communities across seasons,” *Sci. Total Environ.*, vol. 759, 2021, Art. no. 143430.
[4] S. Smoliński and F. Berg, “Varying relationships between fish length and scale size under changing environmental conditions–multidecadal perspective in Atlantic herring,” *Ecological Indicators*, vol. 134, 2022, Art. no. 108494.

[5] E. Maire et al., "Micronutrient supply from global marine fisheries under climate change and overfishing," *Curr. Biol.*, vol. 31, no. 18, pp. 4132–4138, 2021.
[6] X. Xu, W. Li, and Q. Duan, "Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification," *Comput. Electron. Agriculture*, vol. 180, 2021, Art. no. 105878.
[7] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, "Deepfish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing*, vol. 187, pp. 49–58, 2016.
[8] E. Prasetyo, N. Suciati, and C. Fatichah, "Multi-level residual network VGGNet for fish species classification," *J. King Saud Univ.- Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5286–5295, 2022.
[9] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, Jul. 2013.
[10] A. A. Muksit, F. Hasan, M. F. H. B. Emon, M. R. Haque, A. R. Anwary, and S. Shatabda, "Yolo-fish: A robust fish detection model to detect fish in realistic underwater environment," *Ecological Inform.*, vol. 72, 2022, Art. no. 101847.
[11] H. Yang, K. Zheng, and J. Li, "Open set recognition of underwater acoustic targets based on GRU-CAE collaborative deep learning network," *Appl. Acoust.*, vol. 193, 2022, Art. no. 108774.
[12] L. Li, F. Shi, and C. Wang, "Fish image recognition method based on multi-layer feature fusion convolutional network," *Ecological Inform.*, vol. 72, 2022, Art. no. 101873.
[13] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.
[14] Y. Lyu, L. Jing, J. Wang, M. Guo, X. Wang, and J. Yu, "Siamese transformer with hierarchical concept embedding for fine-grained image recognition," *Sci. China Inf. Sci.*, vol. 66, no. 3, 2023, Art. no. 132107.
[15] W. Jia, Q. Ren, Y. Zhao, S. Li, H. Min, and Y. Chen, "EEPNet: An efficient and effective convolutional neural network for palmprint recognition," *Pattern Recognit. Lett.*, vol. 159, pp. 140–149, 2022.
[16] X. Geng, G. Dong, Z. Xia, and H. Liu, "SAR target recognition via random sampling combination in open-world environments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, no. 12, pp. 331–343, Dec. 2022.
[17] M. Liu, Z. Chai, H. Deng, and R. Liu, "A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, no. 5, pp. 4297–4306, May 2022.
[18] G. Cheng, P. Lai, D. Gao, and J. Han, "Class attention network for image recognition," *Sci. China Inf. Sci.*, vol. 66, no. 3, 2023, Art. no. 132105.
[19] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1563–1572.
[20] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4016–4025.
[21] L. Neal, M. Olson, X. Fern, W.-K. Wong, and F. Li, "Open set learning with counterfactual images," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 613–628.
[22] L. Lu, P. Wang, and Y. Cao, "A novel part-level feature extraction method for fine-grained vehicle recognition," *Pattern Recognit.*, vol. 131, 2022, Art. no. 108869.
[23] H. Tang, C. Yuan, Z. Li, and J. Tang, "Learning attention-guided pyramidal features for few-shot fine-grained recognition," *Pattern Recognit.*, vol. 130, 2022, Art. no. 108792.
[24] T. Yan, J. Shi, H. Li, Z. Luo, and Z. Wang, "Discriminative information restoration and extraction for weakly supervised low-resolution fine-grained image recognition," *Pattern Recognit.*, vol. 127, 2022, Art. no. 108629.
[25] D. Wang, Z. Shen, J. Shao, W. Zhang, X. Xue, and Z. Zhang, "Multiple granularity descriptors for fine-grained categorization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2399–2406.
[26] J. Wang, N. Li, Z. Luo, Z. Zhong, and S. Li, "High-order-interaction for weakly supervised fine-grained visual categorization," *Neurocomputing*, vol. 464, pp. 27–36, 2021.
[27] Y. Du, T. Rui, H. Li, C. Yang, and D. Wang, "DeepBP: A bilinear model integrating multi-order statistics for fine-grained recognition," *Comput. Elect. Eng.*, vol. 105, 2023, Art. no. 108432.
[28] B. Huang, G. Chen, H. Zhang, G. Hou, and M. Radenkovic, "Instant deep sea debris detection for maneuverable underwater machines to build sustainable ocean using deep neural network," *Sci. Total Environ.*, vol. 878, 2023, Art. no. 162826.
[29] N. Zhao, B. Huang, J. Yang, M. Radenkovic, and G. Chen, "Oceanic eddy identification using pyramid split attention u-net with remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, no. 2, pp. 1–5, Feb. 2023.
[30] H. Zhang, B. Huang, G. Chen, L. Ge, and M. Radenkovic, "An efficient oceanic eddy identification method with XBT data using transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, no. 11, pp. 9860–9872, Nov. 2022.
[31] P. L. Roberts, J. S. Jaffe, and M. M. Trivedi, "Multiview, broadband acoustic classification of marine fish: A machine learning framework and comparative analysis," *IEEE J. Ocean. Eng.*, vol. 36, no. 1, pp. 90–104, Jan. 2011.
[32] G. Chen, P. Sun, and Y. Shang, "Automatic fish classification system using deep learning," in *Proc. IEEE 29th Int. Conf. Tools Artif. Intell.*, 2017, pp. 24–29.
[33] C. Qiu, S. Zhang, C. Wang, Z. Yu, H. Zheng, and B. Zheng, "Improving transfer learning and squeeze-and-excitation networks for small-scale fine-grained fish image classification," *IEEE Access*, vol. 6, pp. 78503–78512, 2018.
[34] P. Zhuang, Y. Wang, and Y. Qiao, "Wildfish: A large benchmark for fish recognition in the wild," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 1301–1309.
[35] A. Akhtarshenas and R. Toosi, "An open-set framework for underwater image classification using autoencoders," *SN Appl. Sci.*, vol. 4, no. 8, pp. 1–10, 2022.
[36] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[38] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
[39] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher, "Supporting ground-truth annotation of image datasets using clustering," in *Proc. IEEE 21st Int. Conf. Pattern Recognit.*, 2012, pp. 1542–1545.
[40] K. Anantharajah et al., "Local inter-session variability modelling for object classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2014, pp. 309–316.
[41] G. Cutter, K. Stierhoff, and J. Zeng, "Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: Labeled fishes in the wild," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops*, 2015, pp. 57–62.
[42] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 499–515.
[43] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Open-set recognition: A good closed-set classifier is all you need," in *Proc. Int. Conf. Learn. Representations*, 2022.
[44] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[45] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," in *Proc. Int. Conf. Learn. Representations*, 2018.
[46] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[47] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *Proc. Euro. Conf. Comput. Vis.*, 2016, pp. 646–661.

Xiaoya Zhang is currently working toward the M.S. degree in computer technology with the School of Computer Science and Technology, Qingdao University, Qingdao, China.

Her research interests include Big Data oceanography and artificial intelligence.

Baoxiang Huang (Member, IEEE) received the B.S. degree in traffic engineering from the Shandong University of Technology, Zibo, China, in 2002, the M.S. degree in mechatronic engineering from Shandong University, Jinan, China, in 2005, and the Ph.D. degree in computer engineering from the Ocean University of China, Qingdao, China, in 2011.

She is currently an Associate Professor with the College of Computer Science and Technology, Qingdao University, Qingdao. Her research interests include remote sensing image processing and analysis, Big Data oceanography, and artificial intelligence.

Ge Chen received the B.S. degree in marine physics, the M.S. degree in satellite oceanography, and the Ph.D. degree in physical oceanography, from the Ocean University of China (OUC), Qingdao, China, in 1988, 1990, and 1993, respectively.

After graduation, he worked as a Postdoctoral Fellow with the French Research Institute for the Exploitation of the Sea, France, during 1994–1996. Since 1997, he has been a Professor of satellite oceanography and meteorology with the OUC. He is the Deputy Dean with the Institute for Advanced Marine Sciences, OUC, and the Chief Scientist for Ocean Science Satellite Missions with the National Laboratory of Ocean Science and Technology, Qingdao. He is the author of more than 100 peer-reviewed scientific papers published in internationally recognized journals. His current research interests include satellite remote sensing of the ocean and Big Data oceanography.

Dr. Chen was the recipient of the National Science Fund for Outstanding Young Scientists awarded by the Natural Science Foundation of China, in 2001, and became the Chair Professor of Cheung Kong Scholars Program nominated by the Chinese Ministry of Education. He served as the Executive Secretary of the International Pan Ocean Remote Sensing Conference (PORSEC) Association during 1998–2002. He was a Member of the Expert Committee on Ocean Technology for the National High Technology Program (the "863" Program) of China nominated by the Chinese Ministry of Science and Technology.

Milena Radenkovic (Member, IEEE) received the Dipl.-Ing. degree in electric and electronic engineering from the University of Nis, Nis, Serbia, in 1998, and the Ph.D. degree in computer science from the University of Nottingham, Nottingham, U.K., in 2002. She is currently an Associate Professor at the University of Nottingham.

She has authored more than 80 papers in premium conference and journal venues. Her research interests include intelligent mobile and disconnection tolerant networking, complex temporal graphs, self-organized security, and distributed predictive analytics with applications to autonomous vehicles, mobile social networks, smart manufacturing, and predictive telemetry.

Dr. Radenkovic was the recipient of multiple EPSRC and EU grants for her research. She has organized and chaired multiple ACM and IEEE conferences and served on many program committees. She is an Editor of premium journals, such as the *Ad Hoc Networks* (Elsevier), IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED COMPUTING, and *ACM Multimedia*.

Guojia Hou received the B.S. degree in computer science and technology from Linyi University, Linyi, China, in 2010, and the Ph.D. degree in computer application technology from the Ocean University of China, Qingdao, China in 2015.

He is currently an Associate Professor with the School of Computer Science and Technology, Qingdao University, Qingdao. His current research interests include underwater vision, image/video processing, and image quality assessment.