

Deep-Sea Debris Identification Using Deep Convolutional Neural Networks

Bing Xue, Baoxiang Huang^{ID}, Member, IEEE, Ge Chen^{ID}, Haitao Li, and Weibo Wei^{ID}

Abstract—Deep-sea debris is a globally growing problem, which is negatively impacting biological and chemical ecosystems. More seriously, the debris is likely to persist in the deep sea for long periods. Fortunately, with the help of the debris detection system the submersibles can clean up the debris. An excellent classifier is critical to the debris detection system. Therefore, the objective of this study is to determine whether deep convolutional neural networks can distinguish the differences of debris and natural deep-sea environment, so as to effectively achieve deep-sea debris identification. First, a real deep-sea debris images dataset is constructed for further classification research based on an online deep-sea debris database owned by the Japan Agency for Marine-Earth Science and Technology. Second, the hybrid Shuffle-Xception network is constructed to classify the deep-sea image as metal, glass, plastic, rubber, fishing net & rope, natural debris, and cloth. Furthermore, five common convolutional neural networks (CNNs) frameworks are also employed to implement the classification process. Finally, the identification experiments are carried out to validate the performance of the proposed methodology. The results demonstrate that the proposed method is superior to the state-of-the-art CNN method and has the potential for deep-sea debris identification.

Index Terms—Channel shuffle, deep convolutional neural network, deep-sea debris identification, deep-sea debris image dataset, group convolution, sea floor.

I. INTRODUCTION

MARINE environment has always been getting more and more attention all over the world. Debris is everywhere from the shallow seas to the open seas, from the coast to seabed [1], [2]. In most cases, these man-made marine trashes will eventually sink to the bottom of the sea [3]. The deep sea is a veritable marine rubbish repository [4]. Deep-sea debris

Manuscript received April 7, 2021; revised July 16, 2021; accepted August 20, 2021. Date of publication August 27, 2021; date of current version September 15, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 42030406, in part by the Marine Science and Technology Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology, Qingdao under Grant 2018SDKJ0102-8, in part by the Ministry of Science and Technology of China under Grant 2019YFD0901001, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2017MD004. (*Corresponding author: Baoxiang Huang*.)

Bing Xue, Baoxiang Huang, and Weibo Wei are with the College of Computer Science, and Technology, Qingdao University, Qingdao 266071, China (e-mail: xvaceb@163.com; hbx3726@163.com; njustwwb@163.com).

Ge Chen is with the School of Marine Technology, Institute for Advanced Ocean Study, Ocean University of China, Qingdao 266100, China (e-mail: gechen@ouc.edu.cn).

Haitao Li is with the College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China (e-mail: taohaili@sina.com).

Digital Object Identifier 10.1109/JSTARS.2021.3107853

causes more serious water pollution and greater damage to the ecological environment than garbage on the sea surface and beach [5]. Submersibles could solve this problem by surveying and picking up submerged marine debris from the seabed with the help of a debris detection system. However, without an excellent classifier, a satisfying detection system is impossible [6]. Consequently, an accurate deep-sea debris classification algorithm is not only essential for the detection system, but also contributes to further scientific research on deep-sea debris and marine ecological protection.

With the popularity of deep learning, some CNNs such as ResNet [7], GoogleNet [8], as well as graph convolution [9], [10], are applied to various object classifications including the classification of underwater objects. However, most research works on classification of underwater objects focus on the classification of underwater fishes [11], [12], corals [13], plankton [14], [15], unexploded ordnance and wrecks [16]–[18]. As far as garbage research is concerned, synthetic aperture radar technology [19] is often used to monitor marine debris. The RGB camera onboard the unmanned aerospace surveillance on a sandy beach is used to acquire marine macro litter images, and k -nearest neighbor, support vector machine, etc., were used as a comparison method to participate in the classification of these images [20]. The remote sensing satellites [21] and drone aerial photography [22] to obtain rubbish on the ocean surface and beaches are also been adopted to achieve efficient garbage detection [23]–[25] and classification with artificial intelligence methods, including random forest [26] and deep learning methods [27]–[29].

So far, very little attention has been paid to the classification of deep-sea debris with deep learning methods. CNN is employed to detect submerged marine debris from forward-looking sonar imagery [30]. Watanabe *et al.* [31] used YOLOV3 to improve garbage detection and debris floating on the ocean surface, undersea life was detected with 69.6% and 77.2% accuracy, respectively. Meanwhile, the view of marine debris is shifted from the ocean surface to the deep ocean [32]. This article first collates a plastic marine debris dataset, and then uses it to train YOLO, faster R-CNN, etc., to detect underwater plastic and separate plastic from other objects. Although the abovementioned papers have systematically discussed the classification of marine garbage, the problem is that in most of the literature including the abovementioned articles, garbage objects used for the classification of marine garbage only exist on the surface of the ocean and the beach. Although papers [31] and [32] discuss the classification and detection of seabed garbage using

CNN, the accuracy of classification or detection is low and cannot be used in practical applications, and Fulton *et al.* [32] only limits the objects of detection to plastic debris on the seabed. As a whole, the systematic study of classification of deep-sea debris using deep convolutional neural networks is extremely rare.

One of the reasons for the above situations is the scarcity of deep-sea debris datasets. Deep-sea garbage data needs to be captured by specialized deep-sea submersibles using sophisticated imaging equipment in the deep sea, which is costly and difficult. This makes deep-sea garbage datasets very scarce and relevant research cannot be carried out. The other reason is serious intraclass variability and interclass similarity of the seabed garbage. The rubbish flows from land to the deep sea, and is squeezed and eroded by sea water all the year round in the deep sea. The original shape and texture of the rubbish have been greatly different, which intensifies the complexity of the characteristics of deep-sea rubbish. Simultaneously, the optical imaging problems in the ocean and the turbid marine environment have caused the characteristics of marine debris to be no longer prominent in the image. And the temporal and spatial distribution of deep-sea debris and camera differences also make marine debris features diverse. These have caused serious intraclass variability and interclass similarity of deep-sea debris [6], which has posed a particular challenge to the identification of deep-sea debris.

Based on the above analysis, this article intends to solve the above problems. The main work and contributions of this article are as follows.

- 1) The deep-sea debris images dataset named DDI dataset is generated in this article. Compared with the marine litter images on the sea surface or the beach and the trash images dataset generated by simulating the underwater environment, DDI dataset is established based on the deep-sea debris database obtained from the real deep sea by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) [33], so the DDI dataset is truly derived from real deep-sea conditions. All experiments in this article are performed on this dataset, which can achieve robustness and suitability for practical applications to some extent.
- 2) Considering the special characteristics of deep-sea debris, a hybrid convolutional neural network called Shuffle-Xception is constructed, which mainly adopts separable convolution and residual connection. Furthermore, the advantages of group convolution and channel shuffle strategies are combined in Shuffle-Xception architecture. The proposed method can improve the classification accuracy of deep-sea debris significantly.
- 3) Five common CNNs are adopted as the state-of-the-art networks to compare with the proposed Shuffle-Xception. The extensive experiments' results demonstrate that the proposed method has strong data fitting ability and can achieve promising classification results.

This study is among the first of that identifies deep-sea debris using deep learning. The remaining part of this article proceeds as follows. Section II introduces our proposed method in detail,

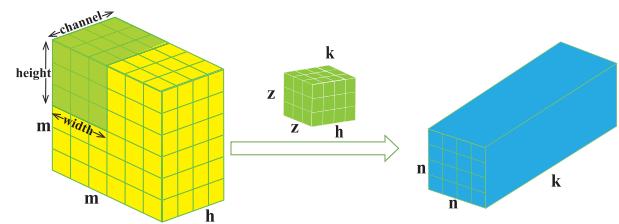


Fig. 1. Ordinary convolution process. m is the size of the feature map to be convolved, and h is the number of channels. z is the size of the convolution filter. k is the number of the filter. n is the size of the feature map after convolution. The ordinary convolution attempts to learn filters in a 3-D space. The 3-D space means 2-D width and height space, and 1-D channel space. This means the convolution simultaneously maps the spatial correlation and channel correlation.

Section III reports the experimental results, and the article concludes with summary in Section IV.

II. METHODOLOGY

Compared with traditional machine learning, many deep convolutional neural networks are more popular in computer vision classification tasks [34]. Taking into account the serious intraclass variability and interclass similarity of the deep-sea debris, this article proposes a new convolutional neural network called Shuffle-Xception based on Xception [35] to achieve accurate identification of deep-sea debris. First, the network mainly adopts separable convolution operations that can extract more abstract and advanced features from the seabed garbage. Second, group convolution strategy is adopted to improve the efficiency and representation ability of the model, meanwhile, channel shuffle strategy is employed to excavate the channel information of deep-sea debris feature map groups thoroughly. Finally, the residual connection is adopted, which can make the network easier to learn, and the information can be transmitted more deeply.

A. Separable Convolution

The structure of depthwise convolution followed by pointwise convolution is called depthwise separable convolution [36]. The main idea of separable convolution is that cross-channels correlations and spatial correlations in the feature maps can be fully decoupled. In ordinary convolution, the convolutional layer learns the information of the spatial dimension of the feature map and the information of the channel dimension at the same time. That is to say, ordinary convolution considers the cross-channels correlations and spatial correlations of the feature map synchronously [35], as displayed in Fig. 1. However, separable convolution assumes that the channel mapping and spatial mapping of the convolution filter can be completely separated. Specifically, the separable convolution operation can be divided into two steps: first, the depthwise convolution is used to complete the separate mapping of each feature map channel; and second, the pointwise convolution (1×1 convolution) operation is adopted to perform the spatial mapping of the feature map.

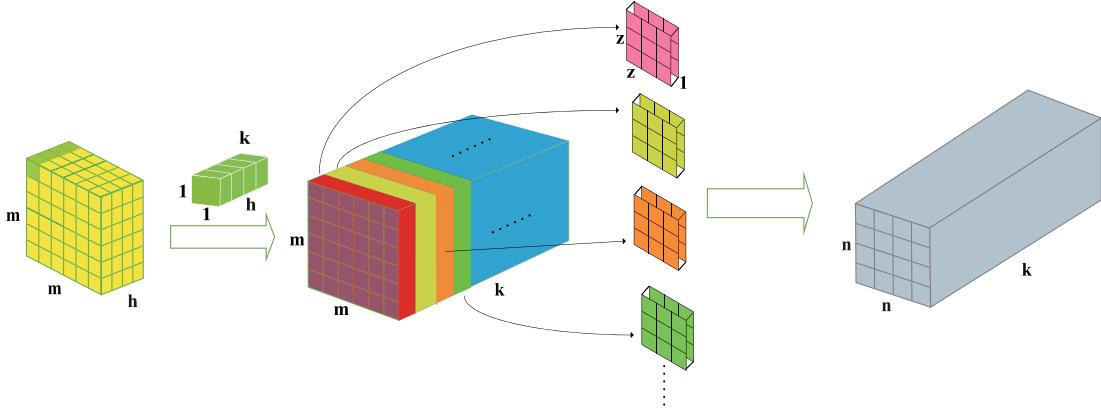


Fig. 2. It first performs 1×1 convolution (pointwise convolution), and then convolves each channel of the feature map separately (depthwise convolution).

The operation can be described as follows:

$$F_j^l = \Gamma_j^{l-1} * \omega_j^l + \alpha_j^l \quad j = 1, 2, \dots, h \quad (1)$$

$$\Gamma_i^{l+1} = \sum_{j=1}^h F_j^l * \mu_{ji}^{l+1} + \beta_i^{l+1} \quad i = 1, \dots, k \quad (2)$$

where Γ_j^{l-1} is the feature map of the j th channel of the $l-1$ th layer; α_j^l is the deviation term; $*$ denotes the convolution operation; ω_j^l is the convolution kernel of depthwise convolution; F_j^l is the feature map of the i th channel of the l th layer; h represents the number of channels of Γ^{l-1} ; β_i^{l+1} denotes the deviation term of other operation. Here, u_{ji}^{l+1} represents the 1×1 convolution kernel of the $l+1$ th layer to convolve with F_j^l . k is the number of channels of Γ^{l+1} .

In fact, the structure of pointwise convolution followed by depthwise convolution as shown in Fig. 2 is equivalent to depthwise separable convolution because these operations are always been used in the stacked setting, the order does not matter much [35]. It is noted that no intermediate activation between pointwise convolution and depthwise convolution in the separable convolution structure used in our proposed method since the activation may bring information loss to the shallow feature space such as depthwise convolution [35].

This convolutional structure makes the cross-channels correlations and spatial correlations mapping of the feature map completely separated, which makes it more powerful to extract high-level abstract features than ordinary convolution [35]. Our proposed Shuffle-Xception network uses this separable convolution to extract the more advanced and abstract category information contained in the seabed debris images, which can accurately determine the category to which the debris belongs. Moreover, this convolution can be more efficient, i.e., it reduces the amount of required parameters and operation costs. For example, Figs. 1 and 2 both illustrate the process of convolving a feature map of size $m \times m$ into a size of $n \times n$. Parameters required for the ordinary convolution shown in Fig. 1 is $z^2 \times h \times k$, while separable convolution in Fig. 2 is only $h \times k + z^2 \times k$. The ratio

of the operation costs of the two convolutions is illustrated as

$$\frac{m^2kh + n^2kz^2}{z^2hkn^2} = \left(\frac{m}{zn}\right)^2 + \frac{1}{h}. \quad (3)$$

The numerator is the operation cost of convolution in Fig. 2, and the denominator is the calculation cost of ordinary convolution. In the actual convolution process, the ratio of this term is always less than 1.

B. Group Convolution and Channel Shuffle

As mentioned in Section II-A, separable convolution uses a combination of pointwise convolution and depthwise convolution so as to realize the spatial mapping and channel mapping of the feature map separately, which can fully mine the information of the feature map. At the same time, through the above analysis, we also know that the number of parameters required for separable convolution is much lower than that of the ordinary convolution. In fact, most of the parameters of separable convolution are mainly provided by pointwise convolution. That is to say, in a separable convolution, the parameters of the pointwise convolution are much more than that of the depthwise convolution, which can be verified from Fig. 2. In Fig. 2, the parameters of pointwise convolution are $h \times k$ and depthwise convolution are $z^2 \times k$. In general, with the deepening of the network depth and the increase in the number of feature map channels, the channel amount h of the point convolution kernel also increases and z^2 is generally a constant ($z = 3$ in the figure). Therefore, the situation $h \times k \gg z^2 \times k$ becomes more obvious as the number of network layers increases, besides, the operation cost $m^2 \times k \times h$ of point convolution is also higher than the operation cost $n^2 \times z^2 \times k$ of depthwise convolution ($m \geq n, h > z^2$). These show that point convolution is the main complexity of separable convolution.

A natural idea is, can full consideration be given to point convolution to make the model more efficient. Group convolution [37] can do this very well. The idea of group convolution: first, the input feature maps are grouped, then these groups are convolved with different filter groups, respectively, and finally the generated feature maps are concatenated together.

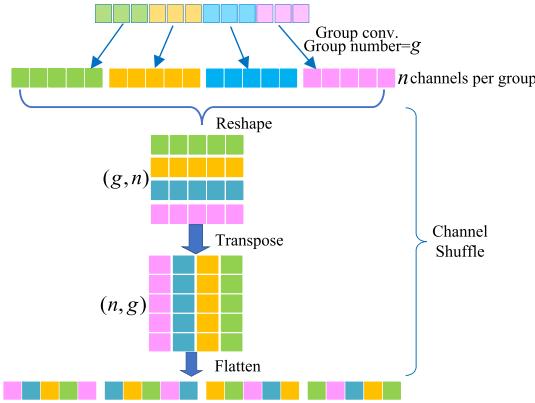


Fig. 3. Group convolution and channel shuffle. The size of the feature map is not considered here, only the channels. Suppose that after group convolution with group number g , we get g groups of feature maps and each group has n channels, then, the g groups of feature maps are reshaped to (g, n) , and transposed to (n, g) , finally flattened. The above operations explain the actual operation process of channel shuffle.

Group convolution greatly reduces the number of parameters, reducing generalization error to a certain extent and making the model more efficient. For example, supposing that the feature map with the number of channels h is mapped to a size of $n \times n$ by a $z \times z \times k$ convolution kernel. As a result, the parameters and operation cost required for ordinary convolution are $z^2 \times h \times k$ and $z^2 \times h \times k \times n^2$, while the group convolution with group number set to g only needs $1/g$ times the parameters and cost of the ordinary convolution. In fact, the filter groups can learn a better representation than 2-D convolution, which is the fundamental reason why the group convolution is effective [38], [39]. Motivated by these, we apply the group convolution to the pointwise convolution in the separable convolution used by Shuffle-Xception.

However, frequently stacking group convolutions will cause the output of a certain group to be only related to the input of that group, blocking the flow of channel information between groups. To solve this problem, channel shuffle operation was introduced [40], as given in Fig. 3. It allows group convolution to obtain input data from different groups, which ensures that the input and output channels are completely correlated and enhances representation.

Group number required for group convolution has an impact on the accuracy of the final classification of deep-sea debris images [39]. Although the use of group convolution can improve the classification ability, it does not mean that the more groups are the better. As group number increases, the channels number in each feature map group decreases, which will cause the information capacity of the deep-sea debris feature map in each group to be too low to support the filter to extract sufficient debris image features, and ultimately weaken representation of the model. A reasonable group number can have a better impact on the classification results. The detailed information will be discussed in Section III-C2.

The structure of the separable convolution combined group convolution and channel shuffle strategies is illustrated in Fig. 4. So, the Shuffle-Xception can not only use separable convolution

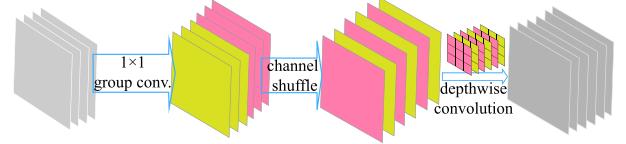


Fig. 4. Main unit/module used by Shuffle-Xception network (SX unit/module).

to extract a large number of rich advanced features from the deep-sea debris feature map, but also employ group convolution to improve the representation level and computational efficiency. Considering the shortcomings of group convolution that the information within the group cannot be shared between groups, channel shuffle is introduced to make the channel information of the deep-sea debris feature map between each group fully correlated.

C. Residual Connection

Intuitively speaking, the deeper network will have the stronger expression ability and the better performance. However, gradient disappearance arises. Consequently, better optimization methods, better initialization strategies, batch normalization (BN) layers, ReLU, and other activation functions have all been used, but their ability to improve the problem is limited. The emergence of residual connections makes deeper network training possible. Shuffle-Xception with residual connections can be described as follows:

$$y^l = \psi(x^l) + H(x^l, \theta^l) \quad (4)$$

$$x^{l+1} = \phi(y^l) \quad (5)$$

where x^l and x^{l+1} represent the input and output of the l th residual unit containing SX modules, respectively; $H(x^l, \theta^l)$ is the residual function, which completes the mapping of SX modules to x^l and θ^l denotes the parameters in the l th residual unit. In Shuffle-Xception, ϕ is identity mapping function and ψ represent identity mapping or 1×1 convolution.

The SX modules combined with the residual connection are shown in Fig. 5. The Shuffle-Xception network can be stacked by several Fig. 5(b) and (c) modules.

D. Network Architecture

The proposed Shuffle-Xception is detailed illustrated in Fig. 6. Generally speaking, the model can be divided into three parts: entry, middle, and exit stage.

In the entry stage, to begin this stage, the image is subjected to ordinary convolution twice for the initial processing of data. Next, an SX unit is performed six times. Among these six times, each twice can be divided into one group, so there are three groups in total, and each group is connected with a shortcut. Convolution filters with a stride size of 2 are used on the shortcut branch to reduce the resolution to facilitate the additional operation. In the middle stage, feature maps are subjected to 24 SX units, and every three units are a group. So, there are eight groups in total. Similarly, each group has an identity shortcut

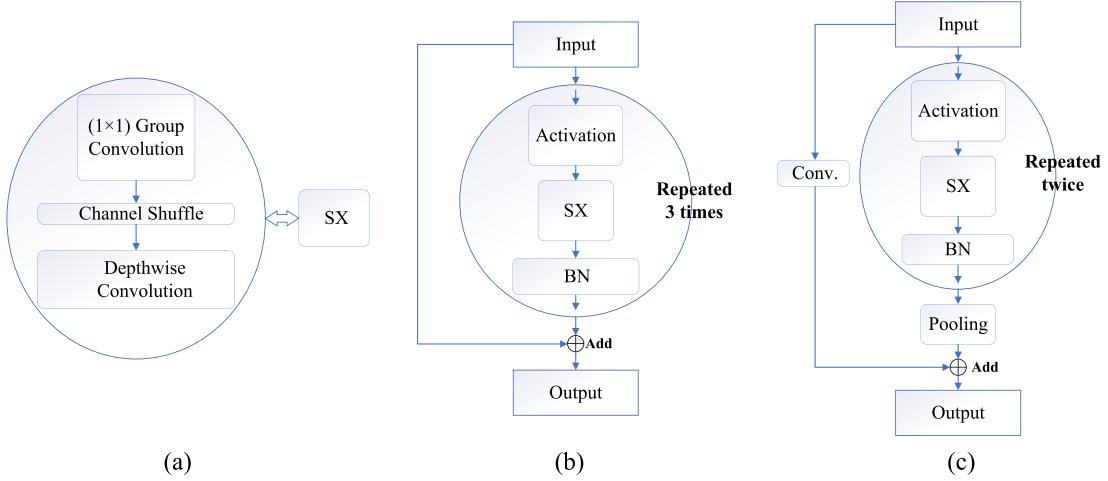


Fig. 5. SX modules combined with the residual connection. (a) gives the SX module. (b) and (c) are the two main modules of the Shuffle-Xception network. ‘Activation-SX-BN’ is repeated three times in module (b), but twice in module (c).

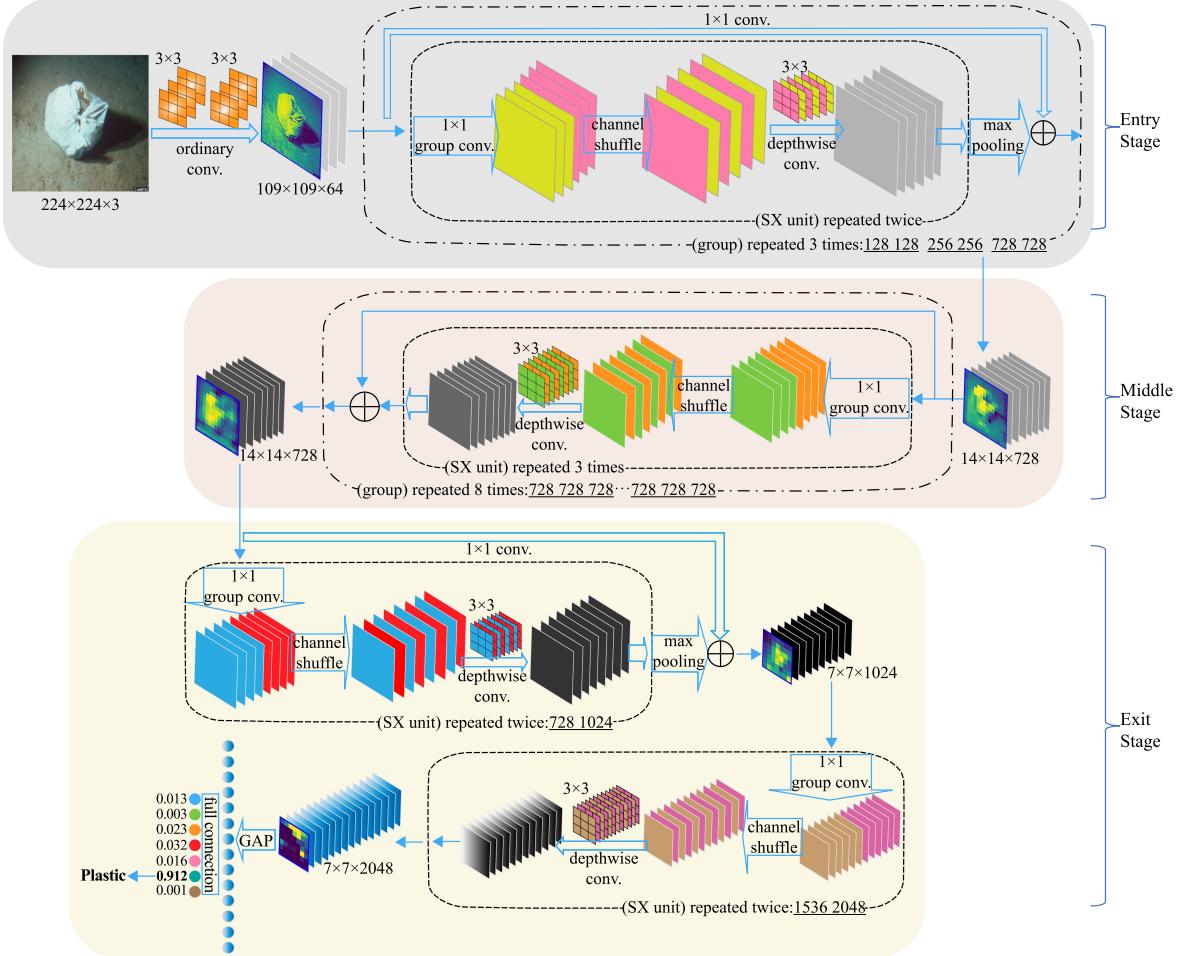


Fig. 6. Structure diagram of Shuffle-Xception based Xception architecture. In the entry stage of the model, the model first uses two ordinary convolutions for the initial processing of data. Subsequently, the following operations called a group are carried out: First, the SX unit is executed twice, and the number of 1×1 group convolution filters in the two SX units is 128 and 128, respectively. Second, the max pooling layer is accessed to reduce the resolution of the feature map. Finally, the obtained feature map and the shortcut branch are added to get the output of this group operation. In the entry stage, the group operation is executed three times repeatedly, and the number of 1×1 group convolution filters of the two SX units in the first group is 128, 128, the second group is 256, 256, and the third group is 728, 728. Parameter description of the middle stage and the exit stage can be deduced according to the figure. The group number required for group convolution in this figure is defaulted to 2.

branch. Note that there is no resolution reduction operation in the branch connection due to the fact that the net do not perform any pooling operation in the residual branch. In the final exit stage, the model performs four SX units. The first two times are a group, and a shortcut connection is also performed. As usual, a convolution operation with a step size of 2 is performed on this connection. The final output layer does not directly fully connect to the output feature maps of the last two SX units for the reason that the huge amount of parameters brought by the full connection may cause overfitting of the model. Accordingly, the model uses global average pooling (GAP) here, which can reduce the number of parameters without affecting the network's ability to classify deep-sea debris images.

The convolution kernels used in the depthwise convolution are all 3×3 . For those convolution filters that perform convolution operations on the shortcut connection, they are all 1×1 in size, and the step size is 2. The network adopts operation sequence shown in Fig. 5(b) and (c) and the activation function is selected as ReLU. The final fully connected layer uses the SoftMax function for probability output for deep-sea debris categories.

III. EXPERIMENTS

A. Dataset Description

Deep-sea debris database is available online as part of the larger JAMSTEC E-Library of Deep-sea Images dataset provided by JAMSTEC and contains photographs and videos of marine debris taken by submersibles “SHINKAI 6500” and “HYPER-DOLPHIN,” in the deep sea [33]. The shooting time of these videos and images that contain deep-sea debris in the database can be traced back to 1982, with a shooting depth of up to 10900 m. This database is entirely composed of debris data on the real seabed, which is different from the marine litter images on the beach, ocean surface, and the debris images generated by simulating the deep-sea environment. It is the database that meets the real conditions in a true sense. So the authenticity and reliability of the deep-sea debris are guaranteed.

JAMSTEC only simply classifies images or videos in the online database, and each video or picture always contains multiple categories of deep-sea debris. Besides, the number of images is relatively fewer compared to the number of videos. The above points indicate that this online database cannot be directly employed for classification tasks. For the sake of constructing a deep-sea debris dataset suitable for image classification, the images and videos from the online database are downloaded and reorganized. To ensure the established dataset can inherit the authenticity and reliability of the database data, we do not do processing that damages the authenticity of the original data so that the classification and identification algorithm trained on this dataset can be implemented in actual scenes as much as possible.

A real deep-sea debris images dataset named DDI is finally constructed, in which each image contains only one type of garbage. It includes 13 914 undersea debris images of seven types of garbage, namely cloth, fishing net & rope, metal, plastic, natural debris, rubber, and glass. All the images are organized from real submerged litter data provided by deep-sea debris database. Table I presents the types of deep-sea debris and the

TABLE I
DEEP-SEA DEBRIS CLASSES AND ITS SAMPLES FROM DDI DATASET

Classes	Plastic	Rubber	Natural debris	Metal	Glass	Fishing net &Rope	Cloth
Samples	2001	1967	1825	2026	1907	2234	1954

amount of debris in each category. Fig. 7 provides some debris images of each category in the dataset. As mentioned before, deep-sea debris has intraclass variability and interclass similarity, which can be seen in Fig. 7. For example, the shape, size, and color of debris within each category are diverse in Fig. 7; the background environment of glasses is changeable; natural debris, glasses, and others have serious visual field dependence; there are interclass similarities between clothing and plastics. For that reason, the feature representation used for our task must be powerful enough to highly represent each major class.

B. Comparative Methods

Five classic deep learning networks that are often used in computer vision classification are selected as the comparative methods. Table II gives their main units, activation function, and whether they use BN operation and regularization. ResNet adopts V2 structure [34] in this article. MobileNet adopts depthwise separable convolution and ReLU6 function that limits the maximum output value of the ReLU function to six [41]. Although both MobileNet and Xception use separable convolution instead of ordinary convolution, an activation function is used between pointwise convolution and depthwise convolution in the separable convolution used by MobileNet, while Xception does not use intermediate activation. The sigmoid function used by the original LeNet [42] is replaced with the popular ReLU, and the subsampling is replaced with the current max pooling of CNN mainstream.

C. Experimental Results

1) *Experiment Settings:* All models used in this experiment are built based on Keras, which runs with TensorFlow, CHTK, and Theano as the backend. Models run on the computer that is equipped with a GeForce GTX 1080Ti GPU and installed with Intel(R) Xeon(R) W-2133 CPU at 3.60 GHz, 31.7 GB RAM, Windows10 operating system.

DDI dataset samples are divided into three subsets: 70% as the training set, 15% as the validation set, and the remaining 15% as the test set. The input size of networks is set to 224×224 commonly used by CNNs, which is the result of tradeoff under the experimental condition. Too small input size may lose information, while too large may cause memory overflow and larger calculations. In addition, with limited memory, if the input image scale is too large, batch size will be restricted (batch size = 1, 2), which may cause the network to obtain an unreasonable classification accuracy. The image is sent to networks after being normalized and adopt data augmentation and early stopping strategies for all networks to prevent overfitting. The learning rate autodecay strategy will be executed if the validation set loss value is not reduced after four epochs; the training will

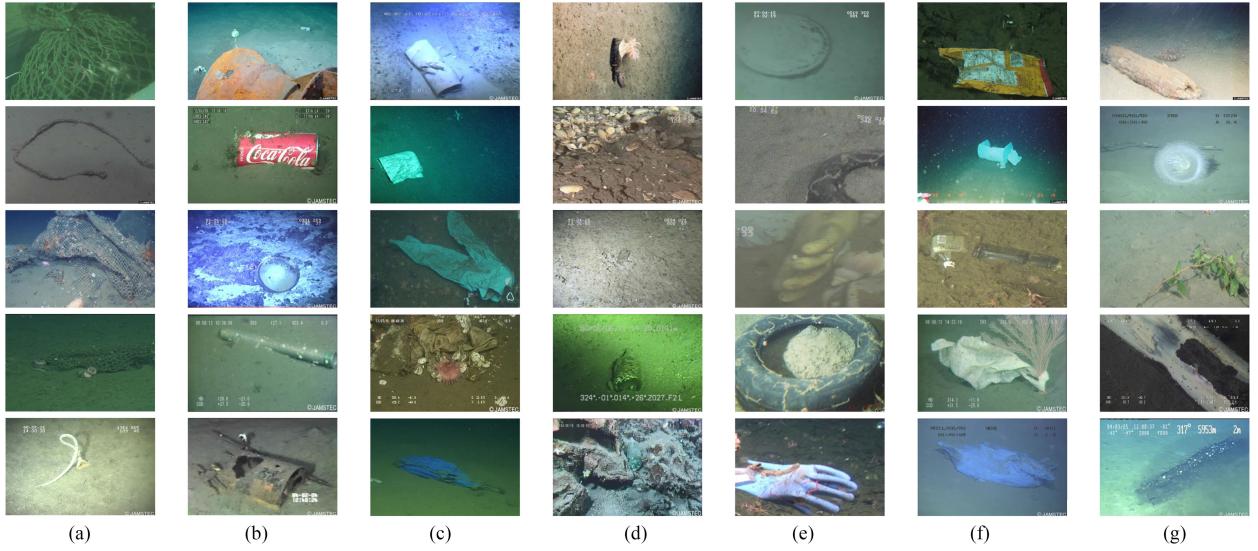


Fig. 7. Examples of seven types of deep-sea debris from DDI dataset. The types of deep-sea debris are fishing net & rope, metal, cloth, glass, rubber, plastic, and natural debris from (a) to (g), respectively.

TABLE II
FIVE COMPARATIVE MODELS AND THEIR MAIN ATTRIBUTES

Model	Main unit	Norm.	Ac. Function	Regularization
ResNetV2 -34/152	Residual unit V2 form, GAP	BN	ReLU	Weight decay
MobileNet	Separable conv., GAP Residual connection	BN	ReLU6	Dropout
Xception	Separable conv., GAP Residual connection	BN	ReLU	-
LeNet	5×5 conv. 2×2 max pooling	-	ReLU	-

GAP represents global average pooling.

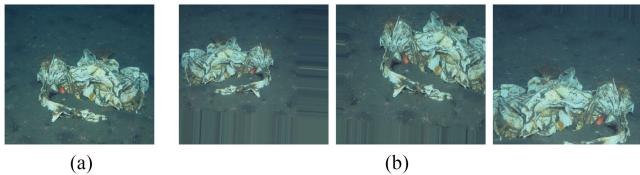


Fig. 8. Examples of deep-sea debris using data augmentation.

be terminated early if the validation set loss value does not decrease after nine epochs. The data augmentation used in the experiments includes random horizontal flip, random zoom in or zoom out, up, down, left, and right translation, and image rotation. Some “fake” debris data can be generated to fill the training set. An example of data augmentation is shown in Fig. 8.

Adam is used as the optimizer. Dropout probability of MobileNet is $1e - 3$ and ResNet uses a weight decay (L2 regularization) rate of $1e - 4$. Shuffle-Xception that does not use other regularization terms is expected to have better results.

For the setting of the learning rate and batch size of models, a technique called the LR Range Test [43] is used to ensure that models find the optimal learning rate in a few iterations. Taking

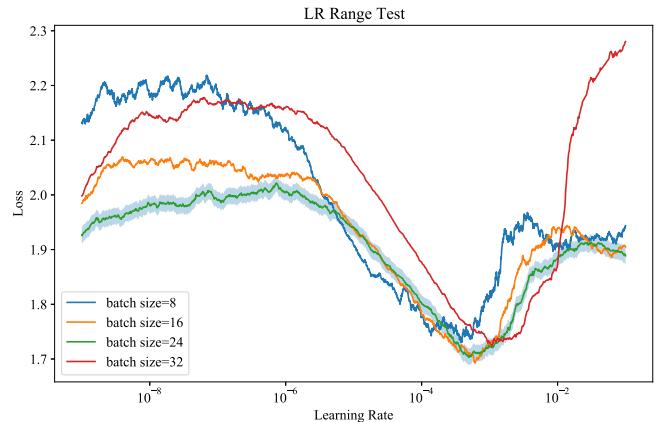


Fig. 9. Curves of LR Range Test with different batch sizes about Shuffle-Xception (*group* = 2).

Shuffle-Xception (*group* = 2) as an example, as depicted in Fig. 9, four experiments with different batch sizes are carried out to find the best learning rate and batch size with the help of LR range test. Closer inspection of Fig. 9, the general trend of all the curves is that the loss value does not change when the learning rate is between 1×10^{-9} and 1×10^{-6} ; the loss value drops rapidly when the learning rate is between 1×10^{-6} and 1×10^{-3} , and the loss value starts to rise when it is greater than 1×10^{-3} . Therefore, 1×10^{-3} is chosen as the initial learning rate of Shuffle-Xception. In addition, it can be seen from Fig. 9 that the larger the batch size, the smoother the curve. This is because that the large batch size can create a stable gradient calculation, which is beneficial to network optimization. Consequently, the batch size of Shuffle-Xception is set as 24 to tradeoff the lower loss value.

2) *Experimental Results and Discussion: Influence of the Group Number.* As mentioned in the previous Section II-B, the group number required for group convolution has an impact on the network representation. The overall classification accuracy

TABLE III
CLASSIFICATION ACCURACY AND SIZE OF SHUFFLE-XCEPTION WITH
DIFFERENT GROUP NUMBERS

Groups	group=1	group=2	group=4	group=8	group=16
OA	0.9149	0.9469	0.9469	0.9364	0.9345
AA	0.9154	0.9470	0.9467	0.9361	0.9343
Parameters	20,875,823	11,117,231	6,229,007	3,784,895	2,686,815

group = 1 Represents that group convolution and channel shuffle strategies are not adopted.

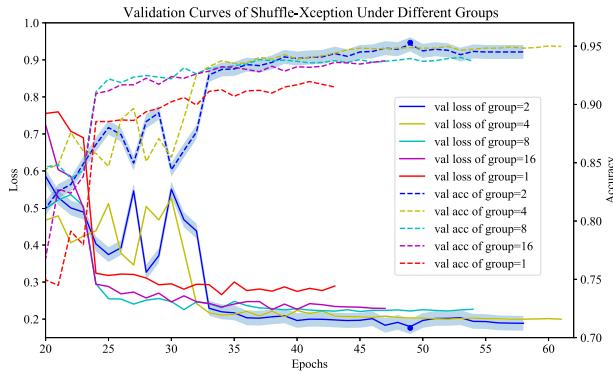


Fig. 10. Validation accuracy and loss curves of Shuffle-Xception with different group numbers. The solid lines represent the loss curves, and the dotted lines represent the accuracy curves. Different colors represent different groups.

(OA) and average accuracy (AA) of Shuffle-Xception under different number of groups are reported in Table III.

Table III implies that group number has a certain impact on the classification accuracy. It can be found that Shuffle-Xception, which uses group convolution and channel shuffle strategies, is always better than unused ones (*group* = 1). Obviously, adopting the strategies will improve the network's ability to identify seabed garbage. This is because the effect brought by the group convolution and the channel shuffle that makes the information fusion between groups allow Shuffle-Xception to obtain information with richer garbage category characteristics. When *group* = 2, 4, the accuracy is the highest, which is because the filter groups at this time can learn the best representation, but as the group number increases, the accuracy decreases. It indicates that excessive reduction of information within each feature map group caused by excessive grouping does have a negative impact on the classification ability of the model. The reason may be that each filter group can not extract sufficient valuable features from the group with too low information content, resulting in information loss.

Fig. 10 intuitively depicts the accuracy and loss changes of the Shuffle-Xception under different group numbers in the actual training process. Although *group* = 1, 8, 16 have the advantage in the early stage, *group* = 2 reaches the best of the whole groups in the 49th epoch of the later stage.

Therefore, the group number of Shuffle-Xception recommended in this article is 2, and the following experiments adopt this setting.

Table III also clearly reports the parameters of the Shuffle-Xception with different groups. Obviously, the number of parameters gradually decreases as the number of groups

increases, which is consistent with the parameter changes caused by group convolution discussed in the previous Section II-B.

Comparative Experiments of All Models. The confusion matrix of all models is built as illustrated in Fig. 11 for the purpose of clearly showing the classification details of each model for each category in the DDI dataset. According to the confusion matrix, kappa values are calculated according to $kappa = (p_o - p_e)/(1 - p_e)$, where p_o is the probability of observed agreement and p_e is the probability of agreement by chance.

In Fig. 11, Shuffle-Xception correctly classifies 284 cloth, 327 fishing net & rope, 285 glass, 276 metal, 259 natural debris, 260 plastic, 290 rubber, and the value on the diagonal of its confusion matrix is the highest among all confusion matrices, which means that it classifies each type of garbage in DDI dataset most accurately among all models. Shuffle-Xception has the highest kappa value, followed by Xception, and LeNet has the worst. Although MobileNet also uses separable convolution, it does not perform well on our deep-sea dataset, which may be due to the intermediate activation between the pointwise convolution and the depthwise convolution in the separable convolution used by MobileNet, which leads to the loss of information [35]. At the same time, the low-precision function ReLU6 used by this model may cause losses to this complex deep-sea debris classification task with high precision requirements. The kappa coefficient of ResNetV2-152 is higher than ResNetV2-34 since ResNetV2-152 is deeper. Even so, Xception, which has a lower number of layers than ResNetV2-152, has achieved a better kappa value due to the use of depthwise separable convolution operations. And our method is higher than the classification result of Xception, which illustrates once again the feasibility and effectiveness of the strategies adopted by Shuffle-Xception.

In the confusion matrix, the color darkens as the number increases. It can be observed from Fig. 11 that the rectangular areas surrounded by metal, natural debris, and plastic of all confusion matrices are always filled with dark colors except for the diagonals. Of course, the color will be relatively lighter with the enhancement of model performance, but this phenomenon still exists. This implies a problem: all models have serious mutual confusion for the three categories of metal, natural debris, and plastic. Similarly, most confusion matrices have similar problems in the lower left corner. As given in Fig. 11(c), we take ResNetV2-34 as an example and mark these areas with red dashed lines.

In order to further demonstrate this phenomenon, a part of the pictures in the red dashed boxes are selected randomly, as given in Fig. 12. It can be inferred from Fig. 12 that the features expressed by deep-sea debris images between different categories have strong similarities (interclass similarity), and at the same time the features expressed by garbage images of the same category also have strong complexity (intraclass variability). These attributes cause most nets to easily confuse these different types of features in the feature extraction process, resulting in a situation where the categories are confused with each other. The similarity and variability are caused by each other. The specific situations where the interclass similarity causes models to confuse garbage classes were discussed in detail in the next paragraph.

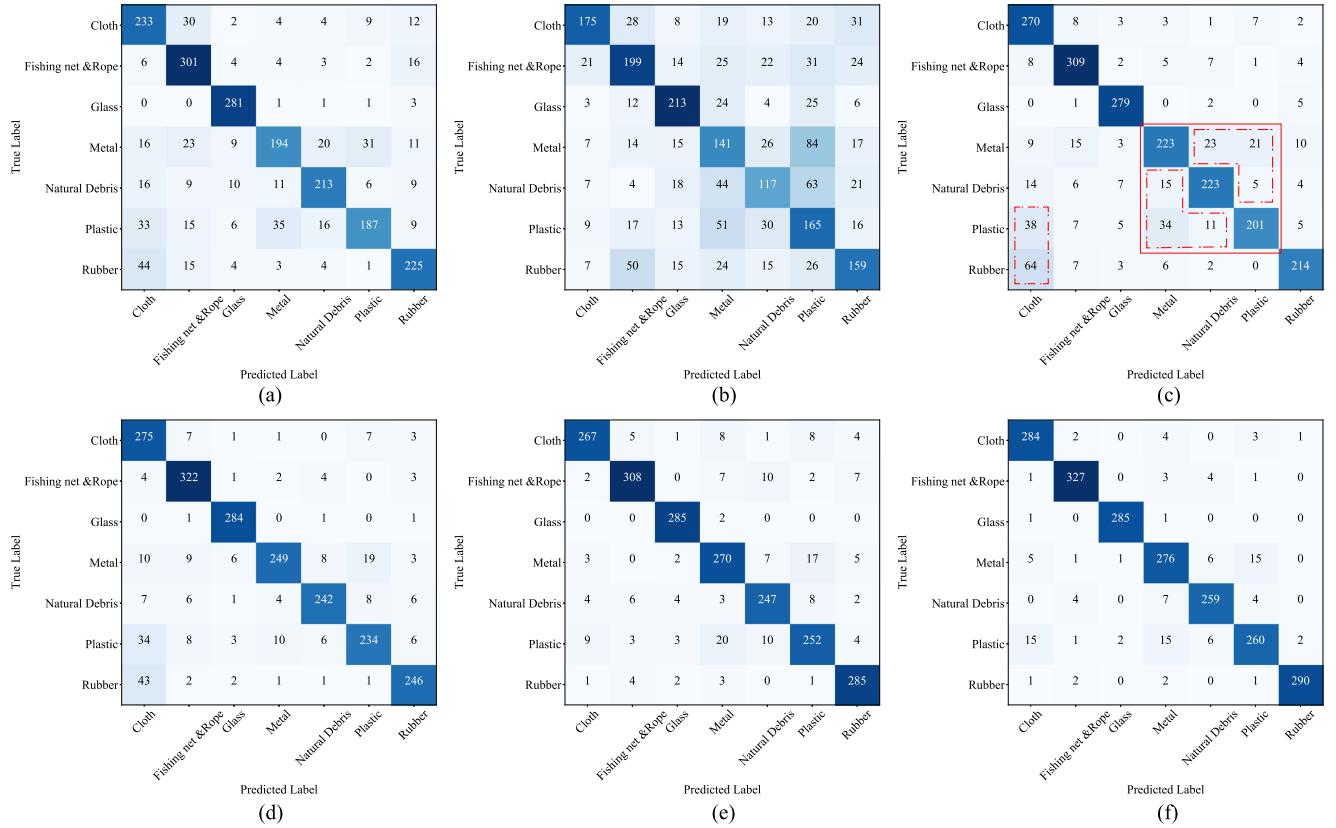


Fig. 11. Confusion matrix of each model on the DDI dataset.

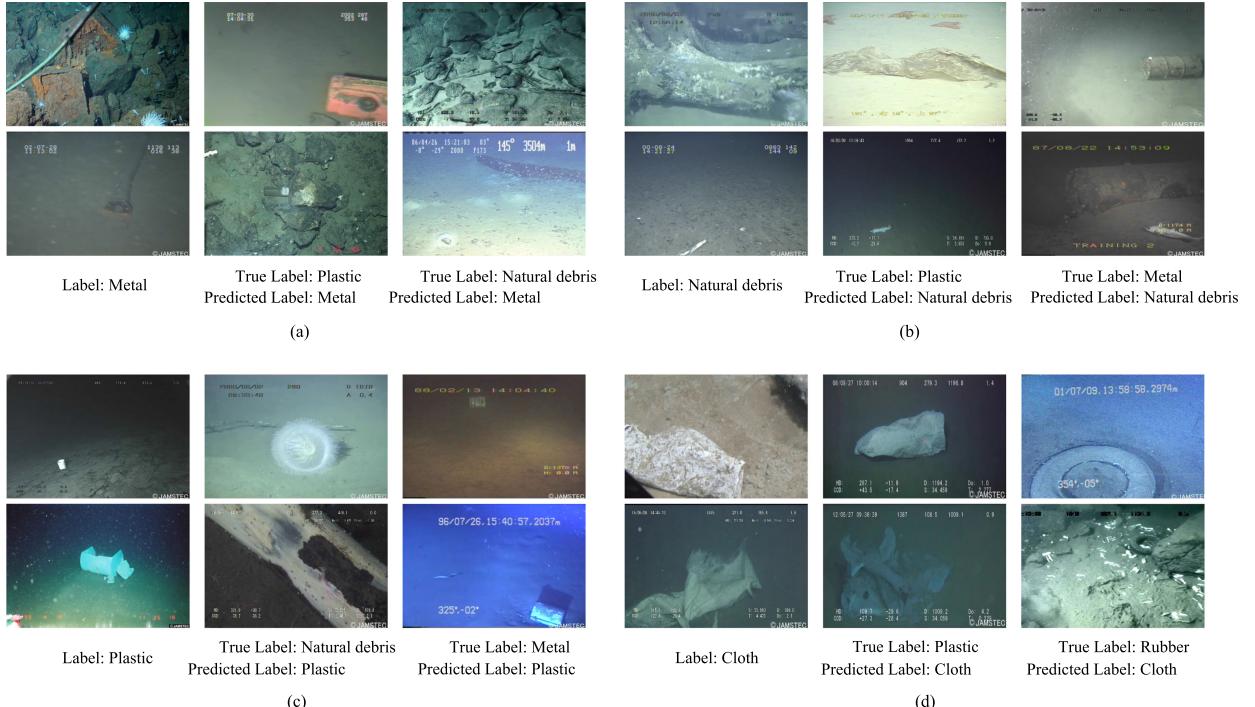


Fig. 12. Some deep-sea debris pictures that are confused by most network models from the red dashed box of all confusion matrices. (a), (b), (c), (d), respectively, shows a category of garbage images and the other two categories of garbage images that are confused into that category.

For intercategory similarity, one is the similarity in appearance between categories. For example, metal buckets and the slender plastics in Fig. 12(b) have strong similarities with natural debris in terms of shape, color, and texture; in Fig. 12(d), the plastic and the cloth are so similar that our eyes cannot clearly distinguish them, and the tires with white attachments on the surface are also recognized as white cloth by most models; the metal bucket and plastic bucket in Fig. 12(c) are almost the same in shape or color, and their smaller size in the images also makes the model unable to clarify the difference between them; similarly, the white trunks in Fig. 12(c) are very similar to plastic buckets in shape and color, so they are misjudged as plastic by most models; the plastic garbage and natural debris in Fig. 12(a) are also very similar to metal in shape and color, which makes most models think that the yellow matrix plastic toy car is a matrix metal showing yellow rust spots, and the black tree trunk is a metal stick. The second is that the deep-sea debris images have a strong viewpoint dependence, which also causes the similarity between different debris categories to a certain extent. For example, there are too many white shells in the background of the second image in the rubber category in Fig. 12(d), so that the models mistakenly regard the white shells as the main body and classify the image as the cloth category, ignoring the small size rubber tires; in Fig. 12(c), since the first image of the natural object category has white marine life, most models consider it to be a plastic, ignoring the tiny branches behind the marine life; the same problem also exists in the plastic bottles and branches in Fig. 12(a); since the plastic bottles and branches occupy a small size in the image, the background is occupied by rocks that look like metal (or it can be said that the rock background of plastic bottles and branches is the same as the rock background in the real metal category), which makes the models ignore the existence of the subject and overconsider the background factors to classify them as metals.

It should be pointed out that except for the two plastic images in Fig. 12(a) that are misclassified as metal by most models containing Shuffle-Xception, the other images that are misclassified by most models are correctly classified by Shuffle-Xception, which to a certain extent means that the Shuffle-Xception network's classification level of deep-sea debris images is higher than other network models. In the next part, heat map method is adopted to focus on showing and explaining the difference in feature extraction capabilities between Shuffle-Xception and other network models when classifying seabed garbage.

A variety of evaluation indicators, including precision as (6), recall as (7), and F value as (8) are adopted to comprehensively evaluate the similarities and differences between all models

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (6)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (7)$$

$$F1_i = \frac{2 \cdot Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (8)$$

where $i = 1, \dots, L$, L is the number of categories. TP_i means the number of correctly classified images of the i th category;

TABLE IV
PRECISION PERFORMANCE OF SIX MODELS FOR EACH CLASS ON THE DDI DATASET

Model	Cloth	Fishing net &Rope	Glass	Metal	Natural debris	Plastic	Rubber
Precision							
ResNetV2-34	0.6700	0.8754	0.9238	0.7797	0.8290	0.8553	0.8770
ResNetV2-152	0.7373	0.9070	0.9530	0.9326	0.9237	0.8699	0.9179
MobileNet	0.6695	0.7659	0.8892	0.7698	0.8161	0.7890	0.7895
LeNet	0.7642	0.6142	0.7196	0.4299	0.5154	0.3986	0.5803
Xception	0.9336	0.9448	0.9596	0.8626	0.8982	0.8750	0.9283
Shuffle-Xception	0.9251	0.9703	0.9896	0.8961	0.9418	0.9155	0.9898
Recall							
ResNetV2-34	0.9184	0.9196	0.9721	0.7336	0.8139	0.6678	0.7230
ResNetV2-152	0.9354	0.9583	0.9895	0.8191	0.8832	0.7774	0.8311
MobileNet	0.7925	0.8958	0.9791	0.6382	0.7774	0.6213	0.7601
LeNet	0.5952	0.5923	0.7422	0.4638	0.4270	0.5482	0.5372
Xception	0.9082	0.9167	0.9930	0.8882	0.9015	0.8372	0.9628
Shuffle-Xception	0.9660	0.9732	0.9930	0.9079	0.9453	0.8638	0.9797
F1							
ResNetV2-34	0.7747	0.8970	0.9474	0.7559	0.8214	0.7500	0.7926
ResNetV2-152	0.8246	0.9320	0.9709	0.8722	0.9030	0.8211	0.8723
MobileNet	0.7259	0.8258	0.9320	0.6978	0.7963	0.6952	0.7745
LeNet	0.6692	0.6030	0.7307	0.4462	0.4671	0.4615	0.5579
Xception	0.9207	0.9305	0.9760	0.8752	0.8998	0.8557	0.9453
Shuffle-Xception	0.9451	0.9718	0.9913	0.9020	0.9435	0.8889	0.9847

FP_i represents the number of images that are incorrectly classified as the i th category; TN_i represents the number of images that do not belong to the i th category and are classified into other categories, and FN_i is the number of the i th category images classified into other classes. $Precision_i$ represents the proportion of true the i th class in the sample predicted to be the i th class. $Recall_i$ refers to the proportion of samples that are actually in the i th class and are predicted to be the i th class. $F1_i$ represents $F1-score$ of the i th category.

The various evaluation index values of all models for each category are given in Table IV. It can be observed that our method is almost optimal in various evaluation indicators, which fully demonstrates the efficiency and feasibility of our method in the task of deep-sea debris classification.

Although Xception achieved the highest $Precision_{cloth}$ with a slight advantage of 0.85% higher than $Precision_{cloth}$ of Shuffle-Xception, it was 5.78% lower than Shuffle-Xception on $Recall_{cloth}$. This indicates that Xception's certainty that samples with the predicted label of cloth really belongs to cloth is slightly higher than that of Shuffle-Xception at the cost of serious omission of the cloth class, which is not cost-effective and leads to the fact that the final classification of the cloth class by Xception is not the best (Low $F1_{cloth}$). Shuffle-Xception dominates the $Precision$ and $Recall$ of each category (except $Precision_{cloth}$), which implies that Shuffle-Xception controls FP and FN of each category to the lowest, which is the result we are most willing to see. It is worth noting that, compared with other models, Shuffle-Xception is more inclined to the situation where $Recall$ is greater than the $Precision$. This illustrates that Shuffle-Xception would rather misreport the debris category than omit the debris, which seems to be more in line with our willingness to clean up deep-sea debris in the actual deep ocean.

The $F1$ values in Table IV take into account both the recall and precision, which can most intuitively show the classification

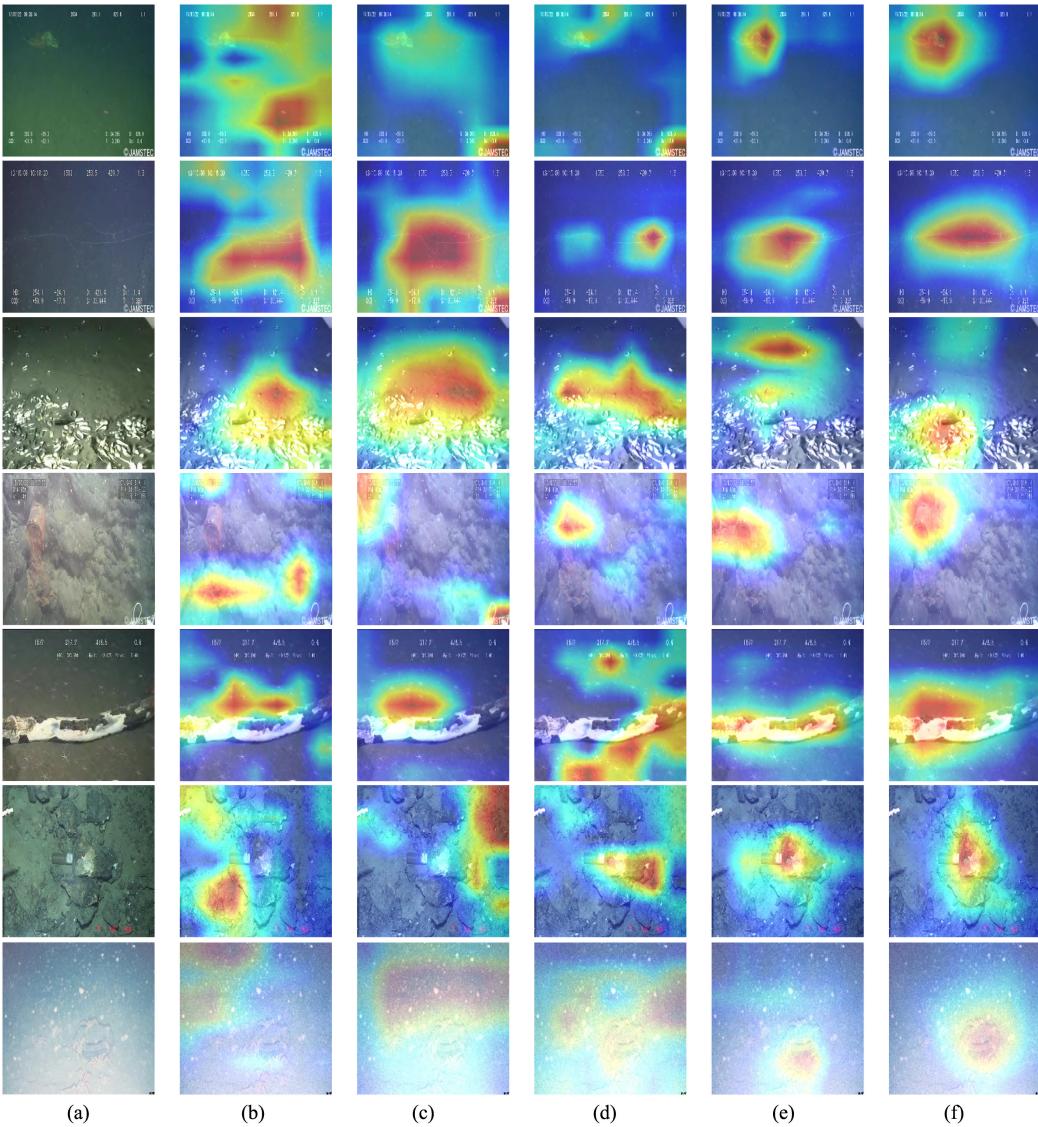


Fig. 13. Area that contributes to the final classification result of the deep-sea debris image is displayed in the form of a heat map. We take the final convolutional layer of each model to form a heat map. Heat map of LeNet is not given here since it has too few layers and poor classification performance. The figure shows a total of seven types of garbage images (a), and heat maps (b)–(f) of five networks for each type of image.

ability of each model for each category. Shuffle-Xception has the best performance on glass, reaching 0.9913, but the worst classification performance on plastic, metal, and natural debris, reaching 0.8889, 0.9020, and 0.9435, respectively. In fact, similar situations exist for most of the remaining models. The reason is that the prominence and uniformity of the characteristics of this kind of garbage caused by the invariance, uniqueness, and antidegradability of glass garbage make the model fit well to this kind of garbage, while most models have low $F1$ values for metal, plastic, and natural debris because of the confusion between the three categories mentioned earlier.

In order to visually give the reasons for the classification differences between models, we deliberately selected pictures with more interference information from each type of deep-sea garbage images as given in Fig. 13. The severe noise interference in the seabed debris image puts a certain pressure on models to

accurately identify the garbage. The heat map reflects the image area that makes a significant contribution to the final classification, which is presented in the form of heat. The contribution of the area to the classification becomes stronger as the heat of the area increases. The classification contribution area reflected by the heat map of Shuffle-Xception is the most accurate and reasonable. The specific location of the garbage is accurately covered by the heat area, highlighting that the debris features of interest extracted by Shuffle-Xception are the most correct and effective, which intuitively explains the reason for the strong classification ability of Shuffle-Xception. The heat areas of other networks do not accurately cover the specific location of the garbage, resulting in the capture of nondebris features, which is detrimental to classification. It can be said that compared with other networks, the feature extraction and analysis capabilities of the Shuffle-Xception occupy an advantage even in a complex environment.

IV. CONCLUSION

This article aims to study classification of deep-sea debris using deep convolutional neural networks. In this article, the first real deep-sea debris images dataset (DDI dataset) is established; Shuffle-Xception network model is proposed and experimented with five other network models. Through the research of this article, the following conclusions can be drawn.

- 1) Constructing a deep-sea debris dataset is crucial for the research of submarine garbage classification, and experiments in this article has demonstrated the feasibility of this dataset in the task of submarine garbage classification.
- 2) The performance of our method on the dataset is satisfactory due to its hybrid architecture. Compared with the five methods, our network is more feasible and efficient in the task of deep-sea debris classification.
- 3) Although plastic, metal, and natural debris are confused with each other, the proposed method gives the best results.

This article provides the first comprehensive assessment of the classification of deep-sea debris using deep learning. In the future, we will explore how to further improve the performance of CNNs in deep-sea debris identification tasks. Furthermore, our work can be combined with object detection methods to locate and mark the location of objects, which is conducive to the further application and implementation of deep-sea detection systems mounted on submersibles.

ACKNOWLEDGMENT

The authors are grateful to JAMSTEC for providing the deep-sea debris database, anonymous reviewers for helpful and constructive comments on the manuscript. The authors would like to thank the anonymous reviewers, theirs comments and perspectives are all valuable and helpful for improving the manuscript greatly, as well as the important guidelines to future research works.

REFERENCES

- [1] I. M. J. van den Beld, B. Guillaumont, L. Menot, C. Bayle, S. Arnaud-Haond, and J.-F. Bourillet, "Marine litter in submarine canyons of the Bay of Biscay," *Deep-Sea Res. Part II-Topical Studies Oceanogr.*, vol. 145, no. SI, pp. 142–152, Nov. 2017.
- [2] F. Galgani and B. Andral, "Methods for evaluating debris on the deep sea floor," in *Proc. Conf. IEEE Oceanic Eng. Soc.*, 1998, pp. 1512–1524.
- [3] S. Chiba *et al.*, "Human footprint in the abyss: 30 year records of deep-sea plastic debris," *Mar. Policy*, vol. 96, pp. 204–212, Oct. 2018.
- [4] W. C. Li, H. F. Tse, and L. Fok, "Plastic waste in the marine environment: A review of sources, occurrence and effects," *Sci. Total Environ.*, vol. 566, pp. 333–349, Oct. 2016.
- [5] K. Schlining *et al.*, "Debris in the deep: Using a 22-year video annotation database to survey marine litter in Monterey Canyon, Central California, USA," *Deep-Sea Res. Part I-Oceanogr. Res. Papers*, vol. 79, pp. 96–105, Sep. 2013.
- [6] M. Valdenegro-Toro, "Deep neural networks for marine debris detection in sonar images," 2019, arXiv preprint *arXiv:1905.05241*. [Online]. Available: <https://arxiv.org/abs/1905.05241>
- [7] 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.
- [8] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [9] S.-H. Wang, V. V. Govindaraj, J. M. Gorri, X. Zhang, and Y.-D. Zhang, "COVID-19 classification by FGNet with deep feature fusion from graph convolutional network and convolutional neural network," *Inf. Fusion*, vol. 67, pp. 208–229, Mar. 2021.
- [10] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Gorri, and S.-H. Wang, "Improved breast cancer classification through combining graph convolutional network and convolutional neural network," *Inf. Process. Manage.*, vol. 58, no. 2, Mar. 2021, Art. no. 102439.
- [11] 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.
- [12] P. L. D. 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.
- [13] A. Mahmood *et al.*, "Deep image representations for coral image classification," *IEEE J. Ocean. Eng.*, vol. 44, no. 1, pp. 121–131, Jan. 2019.
- [14] X. Tang, F. Lin, S. Samson, and A. Remsen, "Binary plankton image classification," *IEEE J. Ocean. Eng.*, vol. 31, no. 3, pp. 728–735, Jul. 2006.
- [15] T. Luo *et al.*, "Recognizing plankton images from the shadow image particle profiling evaluation recorder," *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, vol. 34, no. 4, pp. 1753–1762, Aug. 2004.
- [16] J. J. Hall, M. R. Azimi-Sadjadi, S. G. Kargl, Y. Zhao, and K. L. Williams, "Underwater unexploded ordnance (UXO) classification using a matched subspace classifier with adaptive dictionaries," *IEEE J. Ocean. Eng.*, vol. 44, no. 3, pp. 739–752, Jul. 2019.
- [17] H.-T. Nguyen, E.-H. Lee, and S. Lee, "Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body," *Sensors*, vol. 20, no. 1, Jan. 2020, Art. no. 94.
- [18] G. Huo, Z. Wu, and J. Li, "Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data," *IEEE Access*, vol. 8, pp. 47407–47418, 2020.
- [19] M. Arii, M. Koiwa, and Y. Aoki, "Applicability of SAR to marine debris surveillance after the great East Japan earthquake," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 5, pp. 1729–1744, May 2014.
- [20] G. Goncalves, U. Andriolo, L. Goncalves, P. Sobral, and F. Bessa, "Quantifying marine macro litter abundance on a sandy beach using unmanned aerial systems and object-oriented machine learning methods," *Remote Sens.*, vol. 12, no. 16, Aug. 2020, Art. no. 2599.
- [21] T. Acuna-Ruz *et al.*, "Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications," *Remote Sens. Environ.*, vol. 217, pp. 309–322, Nov. 2018.
- [22] S. W. Jang, S. K. Lee, D.-H. Kim, and Y. Hong-Joo, "The application of unmanned aerial photography for effective monitoring of marine debris," *J. Korean Soc. Mar. Environ. Saf.*, vol. 17, no. 4, pp. 307–314, 2011.
- [23] K. Topouzelis, D. Papageorgiou, A. Karagaitanakis, A. Papakonstantinou, and M. A. Ballesteros, "Plastic litter project 2019: Exploring the detection of floating plastic litter using drones and Sentinel 2 satellite images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 6329–6332.
- [24] K. Themistocleous, C. Papoutsas, S. Michaelides, and D. Hadjimitsis, "Investigating detection of floating plastic litter from space using Sentinel-2 imagery," *Remote Sens.*, vol. 12, no. 16, Aug. 2020, Art. no. 2648.
- [25] K. Topouzelis, A. Papakonstantinou, and S. P. Garaba, "Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 79, pp. 175–183, Jul. 2019.
- [26] G. Goncalves, U. Andriolo, L. Pinto, and D. Duarte, "Mapping marine litter with unmanned aerial systems: A showcase comparison among manual image screening and machine learning techniques," *Mar. Pollut. Bull.*, vol. 155, Jun. 2020, Art. no. 111158.
- [27] J. Lorenzo-Navarro *et al.*, "SMACC: A system for microplastics automatic counting and classification," *IEEE Access*, vol. 8, pp. 25249–25261, 2020.
- [28] L. Fallati, A. Polidori, C. Salvatore, L. Saponari, A. Savini, and P. Galli, "Anthropogenic marine debris assessment with unmanned aerial vehicle imagery and deep learning: A case study along the beaches of the Republic of Maldives," *Sci. Total Environ.*, vol. 693, Nov. 2019, Art. no. 133581.
- [29] S. Kako, S. Morita, and T. Taneda, "Estimation of plastic marine debris volumes on beaches using unmanned aerial vehicles and image processing based on deep learning," *Mar. Pollut. Bull.*, vol. 155, Jun. 2020, Art. no. 111127.
- [30] M. Valdenegro-Toro, "Submerged marine debris detection with autonomous underwater vehicles," in *Proc. Int. Conf. Robot. Autom. Humanitarian Appl.*, 2016, pp. 1–7.

- [31] J.-i. Watanabe, Y. Shao, and N. Miura, "Underwater and airborne monitoring of marine ecosystems and debris," *J. Appl. Remote Sens.*, vol. 13, no. 4, Oct. 2019, Art. no. 044509.
- [32] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," in *Proc. Int. Conf. Robot. Autom.*, 2019, pp. 5752–5758.
- [33] *Deep-Sea Debris Database*, Oct. 2018. [Online]. Available: <http://www.godac.jamstec.go.jp/catalog/dsdebris/e/index.html>
- [34] K. He *et al.*, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.
- [35] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807.
- [36] L. Sifre, "Rigid-motion scattering for image classification," Ph.D. thesis, CMAP, Applied Mathematics Center, Ecole Polytechnique, Palaiseau, France, 2014.
- [37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.
- [38] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5987–5995.
- [39] Y. Ioannou, D. Robertson, R. Cipolla, and A. Criminisi, "Deep roots: Improving CNN efficiency with hierarchical filter groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5977–5986.
- [40] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.
- [41] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," Apr. 2017. *arXiv:1704.04861*. [Online]. Available: <https://arxiv.org/abs/1704.04861>.
- [42] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [43] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2017, pp. 464–472.



Bing Xue is currently working toward the B.S. degree in computer application technology from Qingdao University, Qingdao, China.

His research interests include image processing and analysis, and artificial intelligence.



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

She was an Academic Visitor of the Nottingham University. She is currently an Associate Professor with the College of Computer Science and Technology, Qingdao University, China. 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 all from the Ocean University of China, Shandong, China, in 1988, 1990, and 1993, respectively.

After graduation, he was a Postdoctoral Fellow with IFREMER (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 OUC. He was the Executive Secretary of the International Pan Ocean Remote Sensing Conference (PORSEC) Association during 1998–2002. In 2001, he received the National Science Fund for Outstanding Young Scientists awarded by the Natural Science Foundation of China, and became the Chair Professor of Cheung Kong Scholars Program nominated by the Chinese Ministry of Education. He is currently 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, China. He has authored more than 110 peer-reviewed scientific papers published in internationally recognized journals. His research interests include satellite remote sensing of the ocean and big data oceanography.



Haitao Li received the Graduation degree with a doctorate of science in cartography and geographic information system from the Ocean University of China, Shandong, China, in 2007.

He is currently an Associate Professor with the College of Information Science and Technology, Qingdao University of Science and Technology, China. His research interests include digital ocean, digital environmental protection and security, smart city informatization, geographic information system, positioning and navigation system (Beidou, GPS), mobile Internet of Things, and software engineering.



Weibo Wei received the Ph.D. degree in weapon launch theory and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2006.

He is currently an Associate Professor with the College of Computer Science and Technology, Qingdao University, Qingdao, China. His research interests include image processing and analysis, big data, and artificial intelligence.