

Received March 13, 2021, accepted April 2, 2021, date of publication April 12, 2021, date of current version April 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3072623

Adaptive Selecting and Learning Network and a New Benchmark for Imbalanced Fine-Grained Ship Classification

YIJIE XU¹, MINGHAO YAN², CHENG XU², HUAIDONG ZHANG²,
YIJUN LIU¹, AND XUEMIAO XU^{1,2,3,4,5}, (Member, IEEE)

¹School of Information Engineering, Guangdong University of Technology, Guangzhou 510000, China

²School of Computer Science and Engineering, South China University of Technology, Guangzhou 510000, China

³State Key Laboratory of Subtropical Building Science, Guangzhou 510000, China

⁴Key Laboratory of Big Data and Intelligent Robot, Ministry of Education, Guangzhou 510000, China

⁵Guangdong Provincial Key Laboratory of Computational Intelligence and Cyberspace Information, Guangzhou 510000, China

Corresponding authors: Yijun Liu (yjliu2002@163.com) and Xuemiao Xu (xuemx@scut.edu.cn)

This work was supported in part by the Key-Area Research and Development Program of Guangdong Province, China under Grant 2020B010166003, Grant 2020B010165004, and Grant 2018B010107003; and in part by the NSFC under Grant 61772206, Grant U1611461, and Grant 61472145.

ABSTRACT Ship classification is one of the most essential tasks in ship surveillance, which is an important but challenging problem. Most existing methods are designed for remote sensing images and there are few works processing natural ship images captured by camera. In this paper, we design a new framework called Adaptive Selecting and Learning Network (ASL), to solve the problem of fine-grained classification of ship natural images in the real world. Our method has two key contributions to enable more effective learning from the imbalanced ship data. First, we present a memory network equipped with an adaptive selecting learning strategy to selectively memorize the hard samples that are difficult to classify. The presented learning strategy can re-balance the data distribution of different classes in the training procedure, which achieves more effective learning. Second, we design an inference network with an attention mechanism, to capture the structural similarities between new samples and hard samples. The attention mechanism can enhance the training of hard samples and yields better learning performance. Moreover, our framework can be easily combined with existing fine-grained methods. In addition, we propose a new Dachan Island Ship (DIS) dataset, which is collected in the real-world scenarios. The DIS dataset has a significant imbalanced distribution between classes that aligns with the real situation. Extensive experimental results on the proposed DIS show that our model outperforms most of the existing fine-grained classification methods.

INDEX TERMS Deep learning, imbalanced learning, fine-grained ship classification.

I. INTRODUCTION

In recent years, the success of deep learning [44], [46], [47] makes researchers introduce deep learning technologies into the researches of ship surveillance. As one of the most important tasks in ship surveillance, ship classification has important practical value for many organizations. However, most of the existing researches [57], [58] on ship classification with deep learning technology are specifically designed for hyperspectral [37], [38], infrared thermal [39]–[41], or synthetic aperture radar images [42], which makes them unsuitable for

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

real-world scenarios where we can only get the natural images of ships. The main reasons for the lack of researches on ship classification based on natural images are: 1) The structures of most ships are visually simple, which can only be accurately classified by finding discriminative local details; 2) The data distribution of different categories of ships in real-world scenario is imbalanced; 3) The acquisition of ship natural images is limited by many factors such as scenario, hardware, and weather.

Recognizing the class of a ship accurately is the most significant work in ship surveillance. The existing deep learning networks [30], [36] show great performance in the task of classification due to their superior capability of feature

extraction from images. For ship classification, the discriminative local features should be emphasized as there are only minor differences between different types of ships. To this end, some works [48]–[50] manage to obtain the discriminative local features in the image through adding manual annotations, while some others [14], [51], [52] design various weakly supervised learning mechanisms to capture the discriminative local features of objects when only image-level labels available. Although human-level accuracy of classification can be achieved by these models when applied on the balanced distribution datasets in which the numbers of samples from different classes are generally the same, the performance severely drops when the dataset becomes imbalanced.

In this paper, we propose a new framework that can be used in real-world scenarios named Adaptive Selecting and Learning Network. Without the additional manual bounding-box annotations, we adopt a multi-structure cooperative learning scheme to achieve the effective learning of ship fine-grained expert knowledge under the imbalanced distribution of different categories of ships. Specifically, we use a navigation network to ensure that the model can locate discriminative object parts in the image. When a local region contains more discriminative details of the object, the higher the probability that the region is predicted to belong to the corresponding ground-truth class. By locating these local regions accurately, our model can learn more sophisticated expert knowledge. To accurately classify ships under the imbalanced data distribution of the real world, we design a memory network equipped with an adaptive selecting strategy and an inference network equipped with attention mechanism. According to the learning performance of the model, the memory network adaptive determines which samples of each class of ships contain professional knowledge that will be helpful to improve the performance of the model and store these samples in it. We use the memory mechanism to adaptively re-balance the distribution of different classes in the learning procedure. After obtaining the hard samples that have the similar structure with input, the inference network will use the attention mechanism to accurately capture the relationships between the new sample and hard samples, and then improve the learning of the hard samples while learning the knowledge of new sample.

Since there are few real-world ship datasets of natural images for researches, we propose a new Dachan Island Ship (DIS) dataset which is imbalanced and completely collected in the real-world scenario. Compared with the existing ship classification datasets, the DIS dataset differ from them in three aspects: 1) There are great differences between the samples that belong to the same class due to the shooting angle and illumination condition, while the appearances of different sub-classes are very similar except for some minor differences; 2) There are a larger number of images and types of ships in the DIS compared with the existing datasets, which are expensive and difficult to annotate; 3) The images are all captured in real-world scenarios and the distribution of

different classes is imbalanced. To the best of our knowledge, DIS is the first imbalanced fine-grained ship dataset of natural images.

Our contributions can be summarized in threefold: 1. Delving into the problem of imbalanced fine-grained classification, we propose an Adaptive Selecting and Learning Network, which can adaptive re-balance the distribution of different categories of data according to the learning performance and use the features of the hard samples to optimize the feature of the new sample with attention mechanism. 2. We propose a real-world ship benchmark dataset DIS to advance the frontiers of real-world ship classification research. 3. We conduct extensive experiments on the proposed DIS and the results show that our ASL model performs favorably against the existing fine-grained classification methods.

II. RELATED WORKS

A. FINE-GRAINED CLASSIFICATION

Fine-grained classification aims to perform sub-classification of images belonging to the same basic class. It is more challenging than ordinary image classification tasks due to the subtle inter-class differences and large intra-class differences among those sub-classes. Previous works on fine-grained classification mainly employ complex mathematical methods to extract image features [1]–[4], which are then sent to the multi-stage models to get the label of an image.

In recent years, deep learning has achieved great success in large-scale image classification [30], [36]. Therefore, an increasing number of researches have introduced deep learning techniques to solve fine-grained classification problems [5], [8]. Although discriminative local features can be extracted, these models heavily rely on the annotations of local features or attributes in images, which are expensive and difficult to obtain. To reduce the recline of annotations of local features and attributes, Lin *et al.* propose bilinear pooling [9] and improved it in [10] to obtain more precise feature maps of objects. Jaderberg *et al.* design [7] Spatial Transformer Network which can achieve geometric transformation in the training procedure and consider local parts of objects to improve the accuracy of fine-grained classification. Fu *et al.* [13] design Recurrent Attention CNN that can recursively predict the location of an object through an attention mechanism and combines multi-scale features to predict the class of the object. Zheng *et al.* [14] propose MA-CNN that enables the model to locate multiple parts of an object at the same time. Some recent works, such as [11], introduce meta-learning to facilitate fine-grained classification. Dubey *et al.* [12] combine pairwise confusion loss and cross entropy loss to learn more precise features and guarantee the generalization ability of the model.

Different from the general fine-grained classification, in ship classification, we need to address the issue from heavy imbalanced training samples from different categories. Therefore, we design the learning mechanism, selective learning integrated with the fine-grained feature

extractor borrowed from the fine-grained classification networks, to address the issue from imbalanced learning.

B. IMBALANCED LEARNING

Previous approaches to tackle the problem of imbalanced learning can be categorized into two groups: data resampling and cost-sensitive learning. Data resampling aims to mitigate the imbalanced distribution between classes from the data level. It rebalances the data distribution of different classes in the data preprocessing procedure by resampling the data, which is simple and efficient compared with designing complex algorithms. The typical methods of data resampling include over-sampling minority classes [18], [19], down-sampling majority classes [17], [28], [29] or the combination of over-sampling and down-sampling [15], [16]. Notwithstanding their simplicity, these sampling methods suffer from severe limitations, *e.g.*, over-sampling minority classes will easily cause over-fitting and down-sampling majority classes will lead to the loss of discriminative information in majority classes.

Cost-sensitive learning tries to deal with the imbalanced learning by designing cost-sensitive models at the algorithm level. These methods impose higher penalty cost on the misclassification of minority classes to ensure the models to pay more attention to the learning of minority classes. However, how to design a good cost representation is still a challenging problem. A typical way is to leverage inverse class frequency or pre-defined misclassification costs, which are combined with SVM [20] or decision trees [21]. Boosting is an efficient method proposed by Ting [22]. Considering the limitation of boosting that it is sensitive to noise, Huang *et al.* [23] combine the cost-sensitivity with bagging to obtain a cost-sensitive random forest algorithm. There are also some works combining data sampling or cost-sensitive learning with deep learning technologies to achieve better performance on imbalanced learning [24]–[27]. Different from these methods, we address the imbalanced learning by re-weighting the samples importance with feature selection. The importance of each sample is learned by the proposed ASAL module, so that our framework can automatically re-balance the sample importance by their categories and difficulties on the same time.

C. SHIP SURVEILLANCE

Current ship surveillance researches mainly rely on radar images and thermal infrared images to detect and classify ships. The dataset of synthetic aperture radar images proposed by Wang *et al.* [31] is a representative dataset of radar images. Ships in this dataset are under complex backgrounds, which makes ship detection very challenging. Lang *et al.* [32] propose an improved multi-class adaptive support vector machine combined with image features to achieve transfer learning between automatic identification system data and synthetic aperture radar image data. Sharifzadeh *et al.* [33] design a hybrid CNN-MLP model for SAR ship classification, which processes the image pixel by pixel and uses

the statistical information of its neighbor pixels to detect the target pixels. Wu *et al.* [34] propose a BDA-KELM model that selects the optimal parameters through combining convolutional neural networks, kernel extreme learning machine and dragonfly algorithm in binary space for the classification of high-resolution SAR images. Huang *et al.* [35] design a ship detection model based on multi-scale heterogeneity, which constructs the heterogeneous maps by extracting multi-scale feature maps. The above methods based on deep learning are mainly designed for the SAR images of ships and there are few studies on ship classification or detection using natural images of ships, leading to a barrier for their practical application. To address this issue, this paper focuses on the classification of ships based on natural images and proposes a new dataset derived from real scenarios to facilitate the research on real-world ship classification.

III. SCENARIO AND DATASET

A. DATA ACQUISITION

In this paper, we choose Dachan island in Shenzhen as the sampling scenario. Dachan island, located in the west of the Nanshan district of Shenzhen, China, is the second-largest island of Shenzhen. Geographically, Dachan island is close to the estuary of the pearl river, at the intersection of the National Waterway of Guangdong Province and the Guangdong-Hong Kong Waterway. Since the island is located at the intersection of shipping channels, the waters near Dachan island not only have a large flow of shipping, but also a large variety of ships, which make the surveillance of ships is a huge challenge for customs officers.

Considering that the quality of images taken by cameras after sunset is usually unsatisfactory due to the weak light illumination, we choose to collect data from 9 a.m. to 5 p.m. every day. Images of ships in the data set were collected by surveillance cameras mounted on hills in the central part of the island. The shooting angle of the camera can vary from 30° to 45°, and the illumination is automatically decided by the camera. To ensure that more details of ships can be photographed, we only shoot the ships in the two nearest channels of Dachan island. For each ship, we take several images and just keep the one with the best quality.

B. DACHAN ISLAND SHIP DATASET

DIS contains 13 classes of ships, with a total of 2,500 natural images and the distribution of samples for different classes is extremely imbalanced, which is shown in Table 1. Compared with the existing ship datasets, our DIS has the following properties (see Table 2): 1) DIS is the first fine-grained ship dataset collected in real-world scenarios, which contains images with different clarity and brightness. 2) DIS has a larger number of images, a higher resolution and more fine-grained image-level labels of ships. 3) The data distribution of different classes of ships in DIS is imbalanced, which is more in line with the real-world situation.

Among all the ships, bulk cargo ship, partial container ship, full container ship and sand carrier account for about 70% of the total number of images, while the other nine classes

TABLE 1. Number of samples per category in the proposed DIS dataset.

Class ID	Class	Number
1	Bulk Cargo Ship	346
2	Refrigerator Ship	126
3	Handysize Bulk Carrier	98
4	Partial Container Ship	414
5	Full Container Ship	561
6	Engineering Ship	67
7	Ferry	86
8	Fishing Vessel	59
9	Sand Carrier	493
10	Sand Dredger	164
11	Tug	63
12	Tanker	14
13	Cruise Ship	9
Total		2500

TABLE 2. The statistics comparison between existing ship classification datasets.

Dataset	Images	Classes	Image size
VOC2007	363	1	random
CIFAR-10	6000	1	32×32
Caltech-256	418	4	300×200
DIS	2500	13	1280×720

account for about 30%. The imbalanced distribution of different kinds of ships presents us with a new great challenge in fine-grained classification tasks. We only utilize 11 classes of ships except for tankers and Cruise ships in our fine-grained classification task, since the number of images from the above two is too small.

IV. MODEL

As shown in Figure 2, the proposed Adaptive Selecting and Learning Network (ASL) consists of three components. The feature extractor aims to extract the features of input images. The navigation network captures the discriminative local details of an object and locates these details in the image. The feature vectors of these details and the input image will then be sent into the ASL module. Finally, a probability distribution over all the classes can be obtained. We will discuss each component in detail.

A. NAVIGATION NETWORK

The aim of navigation network is to ensure the model to capture the discriminative local details of the object in the image, which helps the model "see better". The existing fine-grained methods mainly promote classification by two kinds of local features, one is the features of multiple different details of the object, the other is the multi-scale features of one detail of the object. Our ASL module can be combined with these two methods. In this paper, we use Navigator-Teacher-Scrutinizer

**FIGURE 1.** Examples of Dachan Island Ship (DIS) dataset.

Network (NTS) [56] and Recurrent Attention Convolutional Neural Network (RA-CNN) [13] as navigation network to extract local features, respectively. Both of the used networks are designed for finding the local discriminative features to enhance the feature learning for fine-grained classification. Our method can also adopt general features extracted by Resnet only, please refer to the experiments for details.

B. ADAPTIVE SELECTING AND LEARNING MODULE

The proposed ASL module consists of a memory network and an inference network. The memory network adaptively chooses the samples that are difficult to be classified and memorizes them in the learning procedure. By comparing the similarities between new samples and the hard ones in memory, the inference network will choose several hard samples from the memory network to learn from them together with new samples, which enables the network to obtain new knowledge while improving the learning of hard samples. Different from the traditional methods of imbalanced learning, ASL module adaptively adjusts the data distribution between minority and majority classes according to the learning performance of model. We will detail the memory network and inference network, respectively.

1) MEMORY NETWORK

Deep learning models with memory mechanisms are mainly applied to long sequence learning tasks in natural language processing, *e.g.*, RNN. The hidden state of RNN works as a memory of the learned contextual information of the previous time step. However, the memory size of these models are fixed and quite small, which indicates that the model lacks a diversity of the knowledge they have learned. Inspired by the memory network proposed in [53], we propose a memory network with an adaptive selecting strategy that enables the model to automatically choose and memorize hard samples in the training procedure.

2) MEMORY MECHANISM

During the training phase, the memory network stores the hard samples and manages to improve the learning from them along with new samples. The hard samples are stored in the

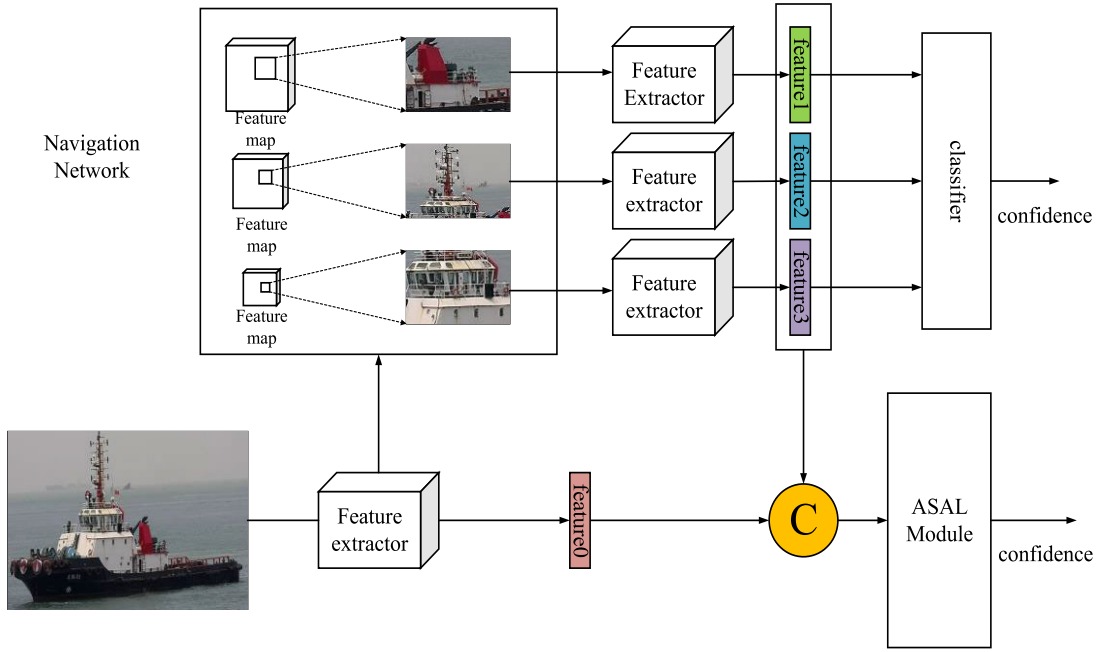


FIGURE 2. The architecture of Adaptive Selecting and Learning Network. The navigation network produces the feature vectors of discriminative local details. The ASL module selects and memorizes the hard samples that are difficult to classify and adaptively rebalance the data distribution of different classes according to the learning performance. “C” means concatenation operation.

form of key-value pairs, which can be formulated as:

$$M = \left\{ \left(m_i^k, m_i^v \right) \right\}_{i=1}^N \quad (1)$$

where N denotes the number of stored samples, m_i^k and m_i^v denote the key and the value of sample x_i , respectively. m_i^v is an integer that denotes the class label of x_i :

$$m_i^k = \text{concatenate} \left(f_v(x_i), f_v(P_1^i), \dots, f_v(P_K^i) \right) \quad (2)$$

where $f_v(x_i)$ denotes the feature vector of the full image, and $\{f_v(P_1^i), \dots, f_v(P_K^i)\}$ denotes the feature vectors of the top- K informative regions of x_i .

Memory network uses Euclidean distance between new sample x_n and difficult samples x_h to calculate similarity weights between samples:

$$w_i = \frac{e^{-d(x_n, x_h^i)}}{\sum_{j=1}^k e^{-d(x_n, x_h^j)}} \quad (3)$$

where $d(x_n, x_h^i)$ is Euclidean distance.

3) ADAPTIVE SELECTING STRATEGY

Whether a sample is hard nor not mainly depends on the prediction made by the model. Given a sample as input, if the model can correctly predict the label of the sample, then the sample is treated as a simple one, otherwise, it is a hard one. We determine whether the sample is a hard one according to the confidence of the ground-truth class predicted by the ASL module. If the confidence is smaller

than a threshold, the sample is thought to be hard and the feature vector of it should be stored in the memory network:

$$U_i = \begin{cases} 1 & \text{if } -\ln(C_{ASAL}(x_i, P_1^i, \dots, P_K^i)) > D(t) \\ 0 & \text{if } -\ln(C_{ASAL}(x_i, P_1^i, \dots, P_K^i)) < D(t) \end{cases} \quad (4)$$

where $C_{ASAL}(x_i, P_1^i, \dots, P_K^i)$ denotes the confidence of sample x_i belonging to ground-truth class, $D(t)$ denotes the selecting function. In this paper, we propose four different selecting functions, which represent four different update strategies of threshold, respectively:

$$D_{line}(t) = \frac{t\varphi}{T}, \quad (5)$$

$$D_{constant}(t) = 0.5\varphi, \quad (6)$$

$$D_{root}(t) = \sqrt{\frac{t}{T}}\varphi \quad (7)$$

$$D_{quadratic}(t) = \left(\frac{t}{T}\right)^2 \varphi \quad (8)$$

where φ , t , and T denote the hyper-parameter, the current epoch, and the total epochs, respectively. For a sample x_i , if $U_i = 1$, its feature vector will be stored in the memory network.

4) INFERENCE NETWORK

Once obtaining the feature vector of the new sample, the memory network will choose its k nearest neighbors. And the feature vector of the new sample is then concatenated with each neighbor, after which, these k updated feature vectors are fed into the inference network to get the label

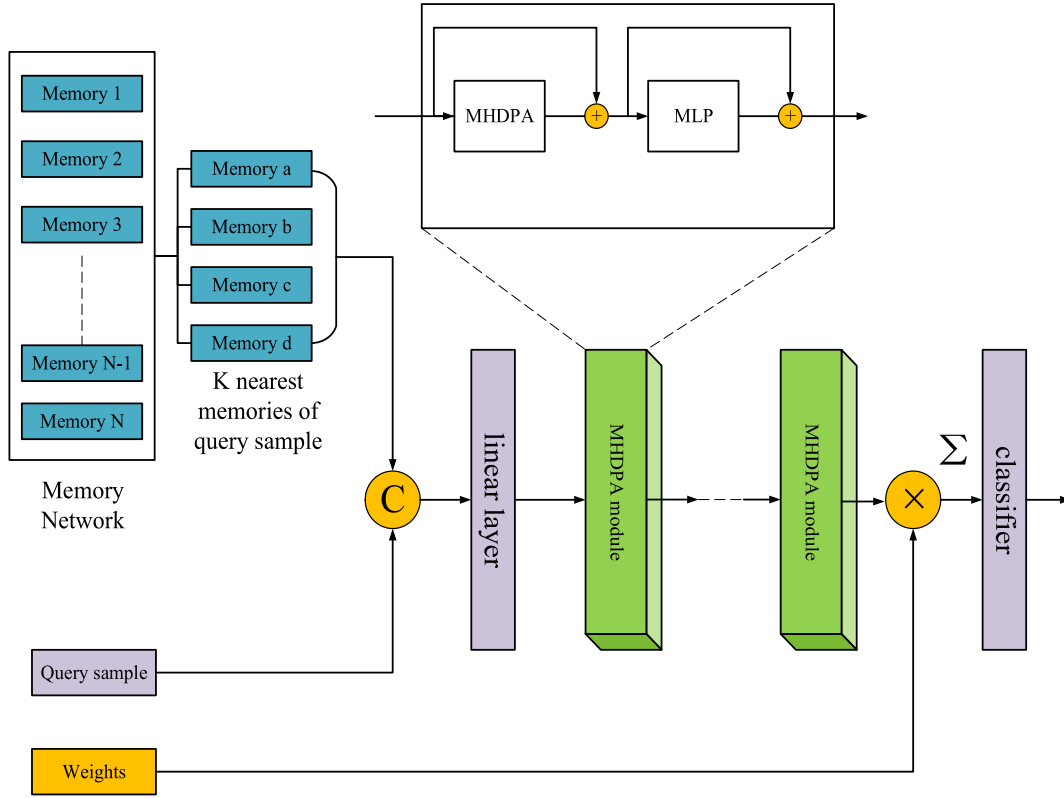


FIGURE 3. The architecture of the ASL module. The weights are calculated by the distance between the query sample and its neighbors. “C” means concatenation operation.

of the new sample. The inference network is composed of a linear layer, N multi-head dot product attention (MHDPA) modules N weight networks and a classifier. The architecture is shown in Figure 3. We adopt multi-head dot product attention (MHDPA) [54] to capture the relationships between a sample and all other memories and update the content of the memory according to the attention information. During the training phase, the MHDPA module not only considers the new samples but also pays attention to the hard samples stored in the memory network.

Specifically, the MHDPA module consists of a multi-head dot product attention layer and a multi-layer perceptron. We store the output of the memory network in a matrix M_K , which has k rows, with row-wise memories. In the MHDPA module, we use the shared weights W^Q , W^K , and W^V to perform linear mapping on the contents of each memory in M_K , by which we get the queries ($Q = M_K W^Q$), keys ($K = M_K W^K$), and values ($V = M_K W^V$). Then we can calculate the dot product of the queries Q and the keys K , which is followed by a Softmax function to obtain a set of weights for calculating a weighted average of values from V :

$$M_n(Q, K, V) = \text{softmax} \left(\frac{M_K W^Q (M_K W^K)^T}{\sqrt{d_k}} \right) M_K W^V \quad (9)$$

where d_k denotes the dimension of the key vectors used as a scaling factor. Finally, we get a new matrix M_n that has the same size as M_k .

A single attention function with keys, values, and queries cannot obtain accurate relationships between memories. By leveraging unique parameters of different shared weights to conduct linear projection on memories in M_k , the MHDPA produces multiple queries, keys, and values sets, based on which, we perform the attention functions in parallel. The outputs of these attention functions are concatenated and projected to generate the final Multi-head attention:

$$\text{MultiHead} = \text{concat}(\text{head}_1, \dots, \text{head}_h) W, \quad (10)$$

$$\text{head}_i = M_n(Q_i, K_i, V_i) \quad (11)$$

Multi-head attention enhances the capability of the model to leverage the information from representation subspace from different positions and boosts the performance of imbalanced fine-grained classification.

We use weight network to optimize the similarity weights between new sample and hard samples in the process of inference:

$$w_l^i = \frac{e^{g_l(MH_l^i, w_{l-1}^i, \theta_l)}}{\sum_{j=1}^k e^{g_l(MH_l^j, w_{l-1}^j, \theta_l)}} \quad (12)$$

where g_l is the l -th weight network, θ_l represents its parameters. MH is the output of MHDPA module. The final feature

is obtained by the weighted sum of the output of the last weight network and the MHDPA module. Then it is sent to the classifier to predict the label of new sample.

C. LOSS FUNCTION

The ASL Network is optimized by two parts of supervisions, the first is the classification loss of ASL module, the other is the loss of the navigation network. Our multi-task loss function can be formulated as follow:

$$L = L_{ASL} + \alpha L_N \quad (13)$$

where L_{ASL} denotes the cross entropy loss of classification. According to different methods of fine-grained feature extraction, we can choose the corresponding loss function of local feature extraction as the loss function of navigation network. In this paper, we use the corresponding loss function of NTS and RA-CNN as L_N , respectively. In our experiment, we set $\alpha = 1$.

V. RESULTS

A. DATASET

To demonstrate the effectiveness of our model, we conduct extensive experiments of imbalanced fine-grained classification on the proposed DIS. We use 11 categories of ships from the dataset with a total number of 2,477 images, among which, 1500 are used as training set, 477 as validation set and 500 as test set. All of the images that we use only have image-level annotations, i.e., the class labels of ships.

B. DEFINITION OF MINORITY AND MAJORITY CLASSES

We define the minority classes of a dataset as

$$\sum_{i \in \text{Class}_{\min}} N_i \leq \rho N \quad (14)$$

where N denotes the total number of images in the dataset and N_i denotes the number of images that belong to class i and ρ is a parameter ranging from 0 to 1. In our DIS dataset, we set $\rho = 0.3$. Therefore, the majority classes consist of four categories of ship and the left seven categories of ships belong to the minority classes. In the following experiments, we evaluate the performance of models through calculating the classification accuracy of majority and minority classes.

C. IMPLEMENTATION DETAILS

We employ ResNet-50 as our feature extractor. The navigation network extracts proposals from the output of the last residual blocks of conv3, conv4, and conv5 of ResNet-50 and selects four local features with the highest confidence, which are fed into the ASL module together with the features of the input image. We use Momentum Stochastic Gradient Descent to train the model with the momentum of 0.9 on two GTX 1080ti GPUs. The initial learning rate is set to 0.001, which is multiplied by 0.1 after 50 epochs. The batch size is set to 16. All the input images used in our experiments are resized to 448×448 .

D. COMPARISON WITH OTHER MODELS

We conduct experiments to demonstrate the superiority of our framework on the proposed DIS dataset. The classification accuracy of our model on DIS are shown in Table 3. The existing methods for fine-grained classification, including two attention-based methods RACNN and MACNN, are presented for comparison. H6-L8 denotes a model consisting of 8 multi-head attention modules and the number of heads in each multi-head attention module is 6. From the experimental results, it can be observed that the accuracy of classification have been improved significantly after the combination of ASL module and two fine-grained methods. Especially, ASL + NTS outperforms these existing methods and achieves 100% classification accuracy in six of the seven minority classes. The performance of ASL + NTS is better than that of ASL + RA-CNN, which indicates that for ship classification, it is more beneficial to extract multiple different detail features than to extract multi-scale features of one detail.

Table 4 summarizes the experimental results of imbalanced learning methods on DIS dataset. All these methods use Resnet-50 as feature extractor, and our ASL module only combines with resnet-50. SMOTE and Down-sampling balance the distribution of classes by different sampling strategies for majority and minority classes. CB-FL and LDAM-DRW optimize the learning of different classes by designing loss function. Compared with these methods, ASL + Resnet has a better performance. This phenomenon demonstrates that ASL module can improve the accuracy of classification by adaptive re-balance the distribution of classes according to the performance of model, and the fusion learning for difficult samples and new samples.

E. ABLATION EXPERIMENTS

Here we analyze the efficacy of the main components of ASL framework. We implement different variants of our framework to analyze their performance on ship classification. The results are shown in Table 5. When we only use Resnet-50 to extract global features for classification, the overall accuracy is only 84.4%. After adding the navigation network and ASAL module respectively, the accuracy of our model on DIS is improved by 8% and 3.8%. When the two modules are adopted together, the average accuracy of the model is 12.2% higher than ResNet-50, which is considered as a significant improvement on our dataset. The experimental results show that our ASL module can improve the accuracy of the model more significantly when the features of targets are more abundant.

To analyze how the update speed of threshold used to determine the hard samples affect the performance of the model in the training procedure, we conduct experiments to evaluate the performances of ASL framework ($H = 6$, $L = 8$) with different selecting functions on DIS dataset. Figure 4 shows the trend of the number of hard samples memorized in the memory network corresponding to different selecting functions. Except the constant function, the square root function has the fastest speed of update before the model

TABLE 3. The classification accuracies of different fine-grained classification methods on DIS dataset.

Model	Cls 1	Cls 2	Cls 3	Cls 4	Cls 5	Cls 6	Cls 7	Cls 8	Cls 9	Cls 10	Cls 11	Min	Max	All
B-CNN [9]	82.5	76.9	60	89.3	92.4	85.7	82.4	91.7	94.9	72.7	84.6	77	90.6	87
FCAN [55]	87.3	76.9	75	88.1	91.6	78.6	88.2	83.3	95.9	75.7	69.2	77.8	91.2	87.6
RA-CNN [13]	81	80.8	70	89.3	93.3	78.6	82.4	100	97.9	81.8	84.6	81.5	91.5	88.8
MA-CNN [14]	88.8	84.6	65	86.9	94.1	100	76.5	91.7	96.9	72.7	84.6	82.2	92.3	89.6
DFL-CNN [45]	95.2	80.8	80	94	89.9	78.6	100	100	100	70	92.3	82.9	94.5	91.4
NTS-NET [56]	84.1	100	70	89.3	99.1	85.7	88.2	100	100	75.7	100	86.7	94.5	92.4
ASL+RA-CNN (H4-L5)	87.3	80.8	75	90.5	95	85.7	94.1	100	98.9	81.8	100	85.9	93.7	91.6
ASL+NTS (H6-L8)	98.4	100	80	91.7	100	100	100	100	100	84.8	100	93.3	97.8	96.6

TABLE 4. The classification accuracies of different imbalanced learning methods on DIS dataset.

Model	Cls 1*	Cls 2	Cls 3	Cls 4*	Cls 5*	Cls 6	Cls 7	Cls 8	Cls 9*	Cls 10	Cls 11	Min	Max	All
SMOTE	82.5	76.9	35	76.2	92.4	64.3	88.2	16.6	86.9	63.6	61.5	60.7	86	79.2
Down-Sampling	79.3	80.8	45	71.4	89.9	71.4	82.4	33.3	84.5	66.7	53.8	64.4	82.4	77.6
CB-FL	88.8	92.3	55	73.8	94.1	78.5	82.4	91.6	88.9	72.7	61.5	71.9	88.7	84.2
LDAM-DRW	90.4	92.3	60	75	95	78.5	88.2	83.3	89.9	75.8	69.2	76.3	89	85.6
ASL+Resnet	93.7	80.7	70	75	93.2	85.7	94.1	91.6	98.9	75.8	84.6	81.4	90.6	88.2

TABLE 5. The classification accuracies of variants with different modules of ASL model.

Model	Cls 1	Cls 2	Cls 3	Cls 4	Cls 5	Cls 6	Cls 7	Cls 8	Cls 9	Cls 10	Cls 11	Min	Max	All
ResNet-50 [36]	90.5	73.1	70	71.4	90.8	85.7	82.4	83.3	92.9	75.7	84.6	75.5	87.7	84.4
ASL (H6-L8)	93.7	80.7	70	75	93.2	85.7	94.1	91.6	98.9	75.8	84.6	81.4	90.6	88.2
NTS	84.1	100	70	89.3	99.1	85.7	88.2	100	100	75.7	100	86.7	94.5	92.4
ASL+NTS (H6-L8)	98.4	100	80	91.7	100	100	100	100	100	84.8	100	93.3	97.8	96.6

TABLE 6. Results of ASL network with different selecting functions on DIS dataset.

Model	Min	Max	All
ASL ($D_{line}(t)$)	88.8	94.7	93.2
ASL ($D_{root}(t)$)	80.7	90.4	87.8
ASL ($D_{constant}(t)$)	91.8	95.6	94.6
ASL ($D_{quadratic}(t)$)	93.3	97.8	96.6

converges, leading to a worse performance of the model than those with other selecting functions. The update speed of the quadratic function is the slowest before the model converges and the corresponding model achieves the highest accuracy. Slower update of threshold and lower threshold allow the model to observe these hard examples more frequently in the training procedure and therefore boosts the performance of the model.

We implement different variants of ASAL module to analyze their performance on DIS. The classification accuracy of these variants are shown in Table 6. When the value of H is fixed ($H = 2$ or $H = 4$) and L increases from 2 to 5, the accuracy achieved by the model for the minority and majority classes classification are improved. However, when

TABLE 7. The classification accuracies of different methods on CUB dataset.

Model	Accuracy
B-CNN [9]	84.1%
FCAN [55]	84.3%
RA-CNN [13]	85.3%
MA-CNN [14]	86.5%
DFL-CNN [45]	87.4%
NTS-NET [56]	87.5%
ASL(H6-L8)	88.4%

L increases from 5 to 8, the performance of the model declines slightly. The results demonstrate that when we fix the value of L and change the value of H, within a certain range, the larger the value of H, the higher the accuracy. The model achieves the highest accuracy when $H = 6$ and $L = 8$. Compared with ASAL model (H4-L5), ASAL model (H6-L8) has a better capability of classifying majority classes.

To verify the effectiveness of ASL framework, experiments are conducted on the public dataset CUB, CIFAR-10 and CIFAR-100. The results of the experiments are summarized in Table 7 and Table 8 respectively. The

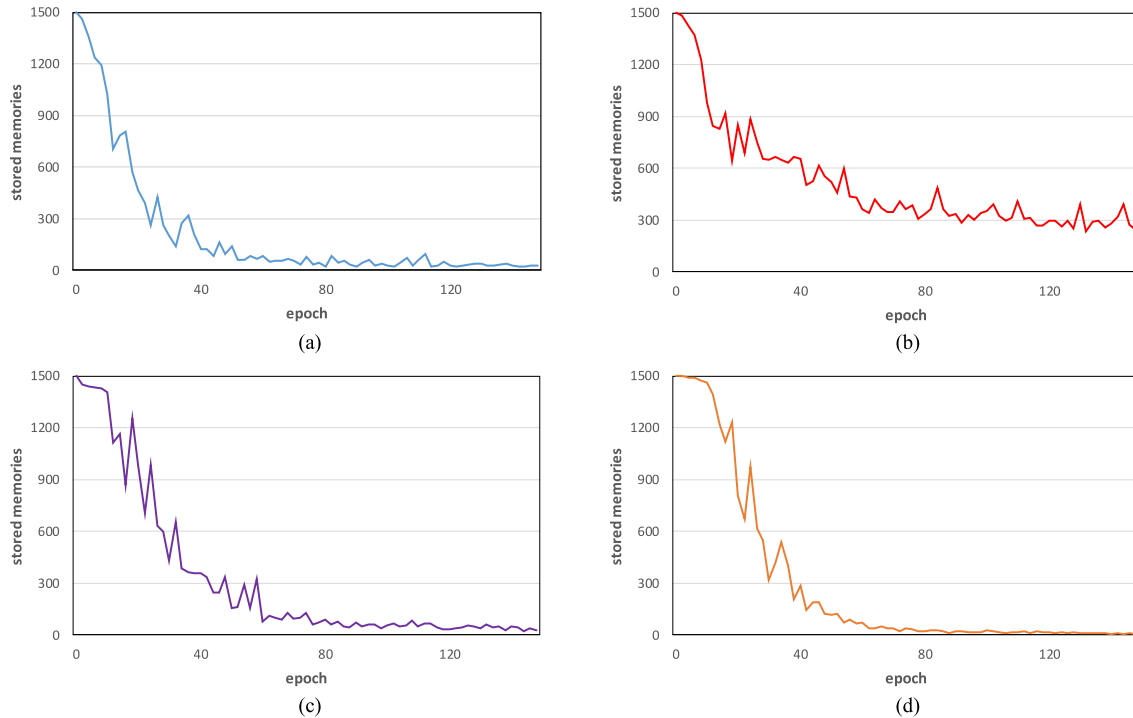


FIGURE 4. The number of samples stored in the memory network. (a), (b), (c) and (d) correspond to the ASL Networks (H6-L8) with constant selecting function, square root selecting function, linear selecting function, and quadratic selecting function respectively.

TABLE 8. The classification accuracies of different methods on CIFAR-10, CIFAR-100 datasets.

Dataset	CIFAR-10		CIFAR-100	
Imbalanced Ratio	100	10	100	10
CB-RS [57]	29.45	13.21	66.56	44.94
CB-FL [57]	25.43	12.90	63.98	42.01
M-DRW [58]	24.94	13.57	59.49	43.78
LDAM-DRW [58]	22.97	11.84	57.96	41.29
ASAL(H6-L8)	21.92	11.42	56.12	40.46

classification accuracy achieved by ASL model on the CUB is 0.9% higher than that of NTS model, which is the best of these existing fine-grained image classification models. We conducted experiments on CIFAR-10 and CIFAR-100 that are the widely used benchmark for imbalanced learning according to the settings in [58]. The top-1 validation errors of various methods for imbalanced CIFAR-10 and CIFAR-100 are reported in Table 8. Our ASAL model performs better than these existing models on these two datasets.

VI. CONCLUSION

In this paper, we propose an Adaptive Selecting and Learning Network (ASL) for adaptive learning from the imbalanced ship data, which aims to fill the gap and spark progress in the real-world ship classification based on natural images. Moreover, we present a new Dachan Island Ship (DIS) dataset with a significant imbalanced distribution between classes. The ASL Network not only locate the discriminative local details

of ships to achieve more accurate classification of ships, but also adaptively re-balance the data distribution in the training procedure and enhance the learning of hard samples while learning new knowledge to achieve a better learning performance from the unbalanced data. Comprehensive experimental results on the proposed DIS demonstrate the superiority of ASL over these existing fine-grained classification methods. In the future, we shall extend the proposed method to the open set problem to better recognize the objects that do not exist in the training data, and integrate the proposed network with the detection network so we can localize each ship if two or more ships are contained in a single image.

ACKNOWLEDGMENT

The authors are grateful for the technical support of customs officers of Dachan Island. (Yujie Xu and Minghao Yan contributed equally to this work.)

REFERENCES

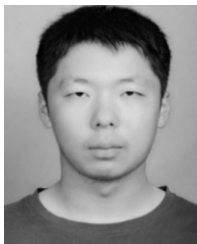
- [1] T. Berg and P. N. Belhumeur, "POOF: Part-based one-vs.-One features for fine-grained categorization, face verification, and attribute estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, Jun. 2013, pp. 955–962.
- [2] L. Yao, Y. Liu, and Y. He, "A novel ship-tracking method for GF-4 satellite sequential images," *Sensors*, vol. 18, no. 7, p. 2007, 2018.
- [3] B. Yao, G. Bradski, and L. Fei-Fei, "A codebook-free and annotation-free approach for fine-grained image categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, Jun. 2012, pp. (CVPR) 3466–3473.
- [4] N. Zhang, R. Farrell, F. Iandola, and T. Darrell, "Deformable part descriptors for fine-grained recognition and attribute prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sydney, NSW, Australia, Dec. 2013, pp. 729–736.

- [5] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based R-CNNs for fine-grained category detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zürich, Switzerland, Sep. 2014, pp. 834–849.
- [6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 580–587.
- [7] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, 2015, pp. 2017–2025.
- [8] D. Lin, X. Shen, C. Lu, and J. Jia, "Deep LAC: Deep localization, alignment and classification for fine-grained recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1666–1674.
- [9] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1449–1457.
- [10] T.-Y. Lin and S. Maji, "Improved bilinear pooling with CNNs," 2017, *arXiv:1707.06772*. [Online]. Available: <http://arxiv.org/abs/1707.06772>
- [11] M. Sun, Y. Yuan, F. Zhou, and E. Ding, "Multi-attention multi-class constraint for fine-grained image recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 805–821.
- [12] A. Dubey, O. Gupta, P. Guo, R. Raskar, R. Farrell, and N. Naik, "Pair-wise confusion for fine-grained visual classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 70–86.
- [13] J. Fu, H. Zheng, and T. Mei, "Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4438–4446.
- [14] H. Zheng, J. Fu, T. Mei, and J. Luo, "Learning multi-attention convolutional neural network for fine-grained image recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5209–5217.
- [15] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Beijing, China, Dec. 2005, pp. 878–887.
- [16] N. Japkowicz, "Learning from imbalanced data sets: A comparison of various strategies," in *Proc. AAAI Workshop Learn. Imbalanced Data Sets*, vol. 68, 2000, pp. 10–15.
- [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
- [18] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
- [19] C. Drummond and R. C. Holte, "C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling," in *Proc. Workshop Learn. Imbalanced Datasets*, vol. 11, 2003, pp. 1–8.
- [20] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 39, no. 1, pp. 281–288, Feb. 2009.
- [21] Z. H. Zhou and X. Y. Liu, "On multi-class cost-sensitive learning," *Comput. Intell.*, vol. 26, no. 3, pp. 232–257, 2010.
- [22] K. M. Ting, "A comparative study of cost-sensitive boosting algorithms," in *Proc. 17th Int. Conf. Mach. Learn. (ICML)*, Stanford, CA, USA, 2000.
- [23] C. Huang, C. C. Loy, and X. Tang, "Discriminative sparse neighbor approximation for imbalanced learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1503–1513, May 2018.
- [24] P. Jeatrakul, K. W. Wong, and C. C. Fung, "Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm," in *Proc. Int. Conf. Neural Inf. Process.* Sydney, NSW, Australia, Nov. 2010, pp. 152–159.
- [25] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan. 2006.
- [26] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3573–3587, Aug. 2018.
- [27] Q. Dong, S. Gong, and X. Zhu, "Class rectification hard mining for imbalanced deep learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 1851–1860.
- [28] T. Maciejewski and J. Stefanowski, "Local neighbourhood extension of SMOTE for mining imbalanced data," in *Proc. IEEE Symp. Comput. Intell. Data Mining (CIDM)*, Paris, France, Apr. 2011, pp. 104–111.
- [29] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 1717–1724.
- [30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.
- [31] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019.
- [32] H. Lang, S. Wu, and Y. Xu, "Ship classification in SAR images improved by AIS knowledge transfer," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 439–443, Mar. 2018.
- [33] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship classification in SAR images using a new hybrid CNN-MLP classifier," *J. Indian Soc. Remote Sens.*, vol. 47, no. 4, pp. 551–562, Apr. 2019.
- [34] J. Wu, Y. Zhu, Z. Wang, Z. Song, X. Liu, W. Wang, Z. Zhang, Y. Yu, Z. Xu, T. Zhang, and J. Zhou, "A novel ship classification approach for high resolution SAR images based on the BDA-KELM classification model," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 6457–6476, Dec. 2017.
- [35] X. Huang, W. Yang, H. Zhang, and G.-S. Xia, "Automatic ship detection in SAR images using multi-scale heterogeneities and an contrario decision," *Remote Sens.*, vol. 7, no. 6, pp. 7695–7711, Jun. 2015.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
- [37] C. González, S. Sánchez, A. Paz, J. Resano, D. Mozos, and A. Plaza, "Use of FPGA or GPU-based architectures for remotely sensed hyperspectral image processing," *Integration*, vol. 46, no. 2, pp. 89–103, Mar. 2013.
- [38] Z. Wang, Q. Yin, H. Li, and B. Hu, "Surface ship target detection in hyperspectral images based on improved variance minimum algorithm," in *Proc. 8th Int. Conf. Digit. Image Process. (ICDIP)*, Chengdu, China, Aug. 2016, pp. 20–22.
- [39] F. S. Leira, T. A. Johansen, and T. I. Fossen, "Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2015, pp. 1–10.
- [40] H.-J. Lee, L.-F. Huang, and Z. Chen, "Multi-frame ship detection and tracking in an infrared image sequence," *Pattern Recognit.*, vol. 23, no. 7, pp. 785–798, Jan. 1990.
- [41] S. Mirghasemi, H. S. Yazdi, and M. Lotfizad, "A target-based color space for sea target detection," *Appl. Intell.*, vol. 36, no. 4, pp. 960–978, 2012.
- [42] J. Jiao, Y. Zhang, H. Sun, X. Yang, X. Gao, W. Hong, K. Fu, and X. Sun, "A densely connected end-to-end neural network for multiscale and multi-scene SAR ship detection," *IEEE Access*, vol. 6, pp. 20881–20892, 2018.
- [43] Z. Liu, F. Zhou, X. Bai, and X. Yu, "Automatic detection of ship target and motion direction in visual images," *Int. J. Electron.*, vol. 100, no. 1, pp. 94–111, Jan. 2013.
- [44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [45] Y. Wang, V. I. Morariu, and L. S. Davis, "Learning a discriminative filter bank within a CNN for fine-grained recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4148–4157.
- [46] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, 2014, pp. 487–495.
- [47] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [48] Y. Chai, V. Lempitsky, and A. Zisserman, "Symbiotic segmentation and part localization for fine-grained categorization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sydney, NSW, Australia, Dec. 2013, pp. 321–328.
- [49] L. Xie, Q. Tian, R. Hong, S. Yan, and B. Zhang, "Hierarchical part matching for fine-grained visual categorization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sydney, NSW, Australia, Dec. 2013, pp. 1641–1648.
- [50] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, "Dog breed classification using part localization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Florence, Italy, 2012, pp. 172–185.
- [51] 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. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 2399–2406.

- [52] B. Zhao, X. Wu, J. Feng, Q. Peng, and S. Yan, "Diversified visual attention networks for fine-grained object classification," *IEEE Trans. Multimedia*, vol. 19, no. 6, pp. 1245–1256, Jun. 2017.
- [53] J. Weston, S. Chopra, and A. Bordes, "Memory networks," 2014, *arXiv:1410.3916*. [Online]. Available: <http://arxiv.org/abs/1410.3916>
- [54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 5998–6008.
- [55] X. Liu, T. Xia, J. Wang, Y. Yang, F. Zhou, and Y. Lin, "Fully convolutional attention networks for fine-grained recognition," 2016, *arXiv:1603.06765*. [Online]. Available: <http://arxiv.org/abs/1603.06765>
- [56] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, "Learning to navigate for fine-grained classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 420–435.
- [57] J. Wang, C. Lu, and W. Jiang, "Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression," *Sensors*, vol. 18, no. 9, p. 2851, Aug. 2018.
- [58] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution SAR images using deep learning of small datasets," *Sensors*, vol. 18, no. 9, p. 2929, Sep. 2018.



YUJIE XU is currently pursuing the bachelor's degree in information engineering with the Guangdong University of Technology. His research interests include computer vision, deep learning, and image processing.



MINGHAO YAN received the B.S. degree from North China Electric Power University, Baoding, China, in 2017, and the M.S. degree from the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, in 2020. His research interests include few-shot learning and computer vision.



CHENG XU received the B.Eng. degree in software engineering from Xiangtan University, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology. His research interests include computer vision, image processing, and deep learning.



HUAI DONG ZHANG received the B.Eng. and Ph.D. degrees in computer science and technology from the South China University of Technology, China, in 2015 and 2020, respectively. He is currently a Postdoctoral Fellow with the School of Nursing, The Hong Kong Polytechnic University, Hong Kong. His research interests include computer vision, image processing, computer graphics, and deep learning.



YIJUN LIU received the B.Sc. degree from Beijing Normal University, China, in 1999, and the M.Phil. and Ph.D. degrees from The University of Manchester, U.K., in 2002 and 2005, respectively, all in computer science and technology. He is currently a Full Professor with the Research Institute of IC Innovation, Guangdong University of Technology, China. His research interests include computer architecture, neuromorphic computing, artificial intelligence, and integrated circuit design.



XUEMIAO XU (Member, IEEE) received the B.S. and M.S. degrees in computer science and engineering from the South China University of Technology, in 2002 and 2005, respectively, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, in 2009. She is currently a Professor with the School of Computer Science and Engineering, South China University of Technology. Her research interests include object detection, tracking, recognition, and image and video understanding and synthesis, particularly their applications in the intelligent transportation.

...