IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2018, accepted June 26, 2018, date of publication July 6, 2018, date of current version July 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2853620

Deep CNN With Multi-Scale Rotation Invariance Features for Ship Classification

QIAOQIAO SHI[1], WEI LI[1], (Senior Member, IEEE), FAN ZHANG[1], WEI HU[1], XU SUN[2], AND LIANRU GAO[2]

[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[2]Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

Corresponding author: Wei Li (liwei089@ieee.org)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0501501 and in part by the Higher Education and High-Quality and World-Class Universities under Grant PY201619.

ABSTRACT With the rapid development of target tracking technology, how to efficiently take advantage of useful information from optical images for ship classification becomes a challenging problem. In this paper, a novel deep learning framework fused with low-level features is proposed. Deep convolutional neural network (CNN) has been popularly used to capture structural information and semantic context because of the ability of learning high-level features; however, lacking of capability to deal with global rotation in large-scale image and losing some important information in bottom layers of the CNN limit its performance in extracting multi-scales rotation invariance features. Comparatively, some classic algorithms, such as Gabor filter or multiple scales completed local binary patterns, can effectively capture low-level texture information. In the proposed framework, low-level features are combined with high-level features obtained by deep CNN. The fused features are further fed into a typical support vector machine classifier. The proposed strategy achieves average accuracy of 98.33% on the BCCT200-RESIZE data and 88.00% on the challenging VAIS data, which demonstrates its superior classification performance when compared with some state-of-the-art methods.

INDEX TERMS Ship classification, convolutional neural network, feature learning, feature-level fusion.

I. INTRODUCTION

Target tracking and classification tasks in the ocean are of great significance for enhancing automatic ship recognition [1] and maritime domain awareness [2]. By using electro-optic sensing signals [3]–[6], the tasks have been widely applied to different fields such as maritime traffic management, vessel traffic services, and naval warfare, maritime search and rescue [7]–[9]. Particularly, with the rapid development of naval strength, ship classification has become more and more important in military and civilian fields [10], [11].

The purpose of ship classification is to identify different categories of ship images with as accurately as possible. Recent years, lots of researchers have made great efforts for this purpose [12], [13]. The visual sensor is usually significantly affected by the natural factors, such as weather and lighting conditions and so on, which makes it difficult to recognize the ship types. What's more, wide within-class variation in some types of vessels also makes ship classification more complicated and challenging [14]–[16]. Up to now, numerous feature extraction algorithms have been proposed for scene classification, including entropy-based hierarchical discriminant regression (E-HDR) [17], completed local binary patterns (CLBP) [18], multi-linears principal component analysis (MPCA) [19], hierarchical multi-scales local binary pattern (HMLBP) [20], histogram of oriented gradients (HOG) [21], and multiple feature learning (MFL) [22]. CLBP is a local texture feature descriptor which captures structure information and extracts local features, e.g., edge feature. Compared with typical LBP, relationship between center point and adjacent points is better described, while the robustness of illumination changes and noise is enhanced. MFL is operated to obtain comprehensive representation of features via fusing the local features and global features. Based on previous work [23], Rainey and Stastny [24] developed several effective recognition algorithms to separate different ship categories.

Gabor filtering was also applied for object recognition, such as face recognition [25], ship classification [26], due to the ability to represent spatial structures of different scales

and orientations. Recently, great attention has been paid to fusion of neural networks and Gabor filter. The benefits of their combination have been confirmed by prior works on hyperspectral image classification [27], human action recognition [28], object recognition in natural scene [29].

Recently, convolutional neural network (CNN) has been successfully employed in many applications, including speech recognition [30], image classification [31], [32], and vehicle type recognition [33], due to its excellent capability of feature representation. Compared with traditional feature-extraction models, deep learning model has capability of learning high-level features from the raw data automatically due to hierarchical architecture [34]. In the field of spatial image analysis, deep CNN architecture has been investigated [35]–[37], and it has been proved to automatically capture the semantic context information from raw data via convolution operation [38]. When it comes to handling the challenge of decreasing between-class difference and increasing within-class difference, the high-level features are more robust and invariant than low-level features.

Although the typical CNN has been widely used in computer version, such as face recognition [39] and scene classification [40], these CNN architectures cannot be applied directly in ship classification tasks. The reasons are explained as follows:

(1) In popular scene classification tasks, the purpose is to separate different kinds of objects in the same picture, such as road and beach, which indicates that the between-class scatter is relatively large. However, for ship classification, the dissimilarity between classes is small. In typical CNN, the first convolution layer may not fully capture all of the local features, resulting in leaving out some important low-level information. Thus, how to make up high-level features produced by the CNN with efficient low-level features to enhance between-class separability is a key issue.

(2) In face recognition tasks, there is no doubt that the face images usually have higher spatial resolution than optimal ship images. Furthermore, there is always a large amount of labeled face data while the number of ship images is much smaller due to the fact that the labeling process is expensive. Hence, good results are obtained when the CNN is employed directly for face recognition. But for ship classification, it may lead to over-fitting or not achieve effective results. How to avoid the over-fitting problem and extract discriminative features from low spatial resolution images is another key point.

In this paper, an efficient feature-fusion learning framework is proposed for ship classification. The proposed method includes two types of features to simultaneously capture structure and local detail information of various ships. After feature extraction, a feature-level fusion strategy integrates different feature vectors into one. The obtained CNN features in the first convolution layer may lead to some important information lost. While Gabor filter can extract information in different directions and edges information, and the MS-CLBP [22] overcomes the limitations of CLBP algorithm in describing the local spatial texture features in a single scale. Both are viewed as the supplementary of high-level feature to obtain more comprehensive one. After that, the fused features are fed into the support vector machine (SVM) classifier [41].

It is always expensive and difficult to collect a large number of labeled training samples for ship classification. Moreover, for deep CNN model, the limitation of the labeled training samples may result in over-fitting. Specifically, it could reflect good classification performance on the training set, but poor on the testing set. The proposed approach that training the CNN through the fine-tuning strategy is able to achieve an excellent performance, even under the condition of insufficient training samples. In the fine-tuning CNN network, such features are extracted from the front layer (near the input image) as color, texture and so on. While the high-level features are captured to the layer closer to the back end, which is more abstract, advanced and specific.

The main contributions of this paper are summarized as follows: (1) The typical CNN is produced by max pooling operation, which is limited to local areas, while global rotation invariance is important for the ship images. In view of this shortcoming, Gabor filter is employed to obtain features in different directions; furthermore, the MS-CLBP operator is used to obtain the local texture, spatial and profile information of ship images to make up for some of the important parts ignored by the typical CNN model. (2) Due to the fact that there are always insufficient labeled ship samples in practice, the designed CNN model is trained by updating weights with fine-tuning strategy, reducing the computational cost. Moreover, Gabor filter and MS-CLBP are unsupervised feature extraction methods that efficient features can be extracted even when the amount of data is very small; that is to say, they are not affected by the number of labeled samples.

The remainder of paper structure is organized as follows. Section II describes the details of proposed classification framework. Section III reports the experimental results and analysis on two datasets (i.e., BCCT200-resize [24] and VAIS [42]). Section IV makes several concluding remarks.

II. PROPOSED CLASSIFICATION FRAMEWORK

Fig. 1 depicts the flowchart of the proposed ship classification framework, including two different features. The first one is that the CNN is employed to capture the high-level features, and the second one is the Gabor-based MS-CLBP for extracting low-level features. Gabor filter is utilized to obtain multiple oriental features, and MS-CLBP is used to extract local contexture. The high-level features are concatenated with the low-level features to obtain a discriminative representation. Finally, fused features are fed into the SVM classifier. Typical SVM has great advantages in solving high dimensional pattern recognition and small sample and nonlinear problems. To a great extent, it can overcome the problems of over-fitting learning and curse of dimensionality. It is noted that the benefits of combining SVM with neural networks have been confirmed by prior work on microvascular morphological

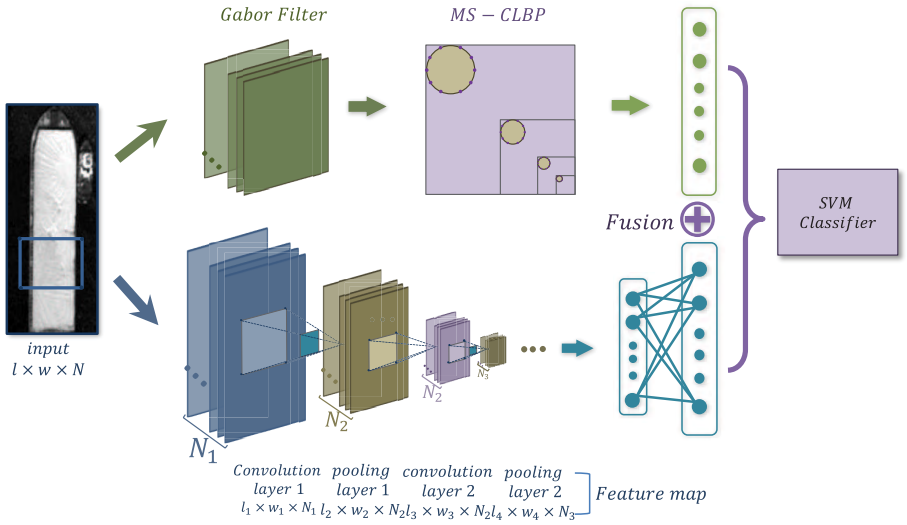

FIGURE 1. A flowchart of the proposed ship classification framework using optical images.

type recognition [43], face recognition [44] and handwritten digit recognition [45].

The motivation of developing the hierarchical architecture of the CNN in ship classification tasks is that the network can learn the high-level features that capture the structure information automatically by convolution operation, which can avoid the complexity of hand-crafted features. Furthermore, in the deep CNN model, low-level features extracted from the first convolution layer are fed into the top layers via layer-to-layer propagation to generate high-level features. Compared with low-level features, they are more abstract and discriminative.

Fig. 2 shows the ship images with different profiles (4 classes selected from BCCT200-resize dataset [24]). The one of each classes has its own shape which is different from other classes, although it is easily affected by some external factors (e.g., wind, cloud, and complicated ocean environments). The fundamental challenge is to model the within-class difference and inter-class similarity. Thus, deep CNN is employed to capture the high-level feature for improving the performance. The key is to select appropriate network structure according to the number of the labeled examples since too many layers may lead to over-fitting in a relatively small-scale ship data. In the view of fact that the number of ship data marked is small, thus the CaffeNet with fewer layers is employed.

A. DETAILS OF THE DESIGNED CNN

The proposed deep CNN architecture is based on CaffeNet, which is a deep convolutional network model trained on the ImageNet data set [46]. Fig. 3 shows details of the deep network which includes eight layers, with five convolutional layers to learn local information and three fully connected layers. Each layer is activated by the rectified linear unit (*ReLU*) function, and most of the noise and secondary

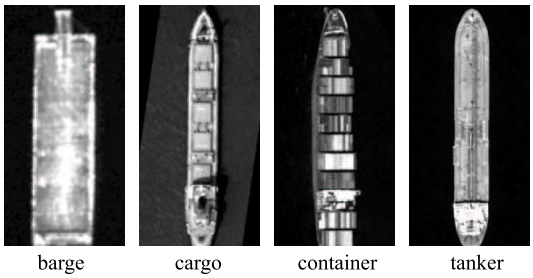

FIGURE 2. Four kinds of ships selected from the BCCT200-resize data set.

features are discarded. The kernel size of convolution layer is set relatively large to accelerate the training process and reduce the number of layers. Since the output size of each node after two levels of convolution is small enough, there is no need to add Max Pooling in the next two layers. In Fig. 3, C represents the number of ship category, and C is 4 for the BCCT200-resize data and 6 for the VAIS data. In order to improve the performance of the network, a strategy of accelerating the training speed of the network (i.e., local response normalization, LRN) is employed in the first two layers. And the Dropout layer is joined in the last two fully-connected layers, improving the generalization ability of the network and avoid over-fitting. Here, the drop parameter is set 0.5.

In the preprocessing phase, we unify the size of the input image to 256 × 256. And in the training phase, the input images would be cut randomly to obtain subregions of the same size 227 × 227, while in the testing phase, the cropping operation is only executed in the center of input images. Experimental results indicate that data augmentation is effective, since the sub-image contains the most important information. Finally, a 4096-dimensional feature vector of the second fully-connected layer is obtained.

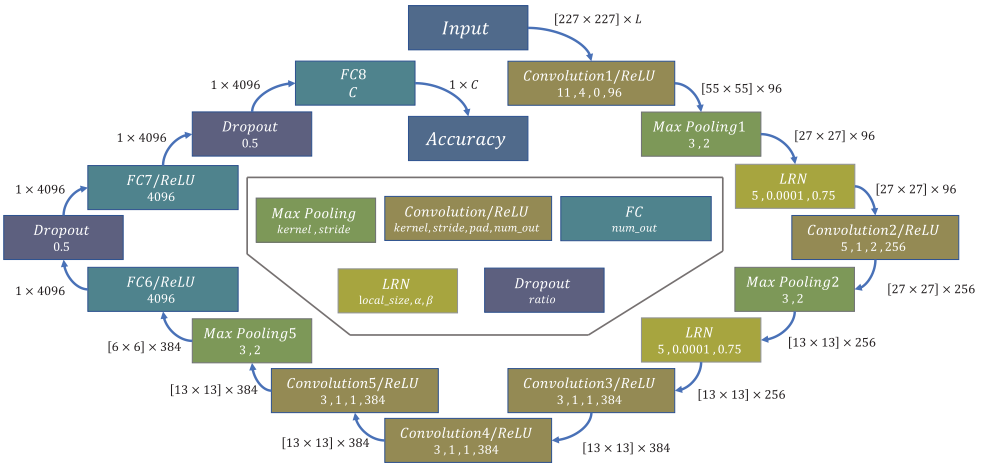

FIGURE 3. Details of the deep network neural model employed in the proposed framework.

Due to that the ship images are cut to 227 × 227, the input size to the first convolution layer of CNN model is 227 × 227 × L. Here, L indicates the channels, and L is 1 for the BCCT200-resize data, and 3 for the VAIS data. The output of the convolution layer can be expressed as $w \times l \times N_i$, where N_i represents the number of neurons in the i layer, and w and l can be calculated using the formula,

$$w = (w_0 + 2 \times pad - kernel\_size)/stride + 1, \quad (1)$$

$$l = (l_0 + 2 \times pad - \mathrm{ker}\,nel\_size)/stride + 1, \quad (2)$$

where w_0 and l_0 represent the width and height of ship image fed into the first convolution layer respectively, *kernel_size* indicates the size of convolution kernel, *pad* represents the edge is expanded with 0 and *stride* is the step length of convolution operation.

It is difficult to train the model up to the optimum from scratch because there is insufficient number of labeled ship images. In order to solve this issue, the fine-tuning strategy is adopted, where the weights from the pre-trained model are significantly reduced. As described in Algo. 1, the training set is firstly prepared well; and then, the mean file of ship images is calculated to accelerate the process of training. The parameters in solver file are also modified to speed up convergence and learning efficiency. Finally, the pre-trained model can provide efficiently high-level features with the same size number of training data.

B. GABOR-BASED MS-CLBP

Gabor filters and MS-CLBP have been developed in the field of computer vision such as scene classification [47], face recognition [48], and vessel type classification [22]. Gabor filter has better directional characteristics to extract features in different directions to enhance global rotation invariance, and MS-CLBP is a better feature descriptor that can capture such vital information of ship as local texture, spatial, profile feature, which overcomes the limitation of sole CLBP in describing the local spatial texture features in a single scale.

Algorithm 1 Fine-Tuning CNN

Require: Prepared training set and testing set

1: Calculate the mean file of training data by using convert_imageset tool under Caffe.
2: Modify the "source" and "meanfile" of the data layer according to LMDB file and mean file generated on the basis of the previous step
3: Modify the name and learning rate of the output layer
4: Reduce the value of "base_lr", "step_size", "test_iter" and "max iteration" in the solver file, and the rest of parameters remains unchanged
5: Load the pre-trained model based on CaffeNet
6: Training

Ensure: High-level features

Thus, the combination of Gabor filter with MS-CLBP is employed to extract low-level features of ship images that as the second type of features in the proposed approach. The features extraction process is described in Algo. 2.

Algorithm 2 The Combination of Gabor Filter With MS-CLBP

Require: Prepared training set and testing set

1: Capture the characteristics of Gabor operator by convolving input ship images
2: Tune parameters (m, r) of MS-CLBP operator and select the optimal parameters
3: Obtain MS-CLBP features based on Gabor feature space for each ship image
4: Produce the final low-level features by cascading the all the MS-CLBP features

Ensure: Low-level features

In the time domain, a Gabor filter can be viewed as the multiplication of the Gaussian function with the sinusoidal plane wave [49]. The Gabor function is composed of

two parts of the real and imaginary parts. The real part smooths the image by filtering, while the imaginary mainly reflects in the image of the edge detection. The result that product of the Gaussian function with a sine wave is the impulse response of Gabor filter. The mathematical expression form of two dimensional Gabor functions,

$$g(x, y; \lambda, \theta, \sigma, \gamma) = \exp(-\frac{x_0^2 + \gamma^2 y_0^2}{2\sigma^2}) \cdot \exp(i(2\pi \frac{x_0}{\lambda} + \psi)) \tag{3}$$

$$\begin{cases} x_0 = x\cos\theta + y\sin\theta \\ y_0 = -x\sin\theta + y\cos\theta \end{cases} \tag{4}$$

where x and y are the location of the pixels in the space, λ is the wavelength (note that its value is usually greater than or equal to 2 but less than 1/5 of the input image), γ is the aspect ratio which determines the ellipticity of the Gabor function (its value is 0.5), θ is the direction that regulates the direction of the parallel stripes when the Gabor function processing the image, taking values between 0 and 360 degrees, and ψ is the phase offset (its value range is from -180 to 180 degrees). Note that $\psi = 0$ and $\psi = \frac{\pi}{2}$ returns the real and imaginary parts of the Gabor filter, respectively. Here, bw is the bandwidth, whose value is related to the ratio of the $\frac{\sigma}{\lambda}$, and σ is the standard deviation with the expression as follows,

$$bw = \log_2 \frac{\frac{\sigma}{\lambda}\pi + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda}\pi - \sqrt{\frac{\ln 2}{2}}} \tag{5}$$

$$\frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2} \cdot \frac{2^{bw}+1}{2^{bw}-1}}. \tag{6}$$

The LBP is used as a local texture feature descriptor to extract the spatial texture features of ship images [22], which has been widely employed for face recognition [50] and texture classification [51]. Given a central point A, whose pixel value is denoted as g_c. The m neighborhood pixels $A_i(i = 0, 1, \cdots, m-1)$ on the circumference of the radius r are compared with the central pixel. The LBP value of the center pixel A is defined as follows,

$$LBP_{m,r}(A) = \sum_{i=0}^{m-1} s(g_i - g_c)2^i$$

$$s(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x \leqslant 0 \end{cases} \tag{7}$$

where r is the distance from the center point A, and m is the number of nearest neighbors, g_i represents the gray value of the neighbors, and $x = (g_i - g_c)$ represents the difference between the center pixel and each neighbor. The CLBP algorithm is improved on the basis of the classic LBP. It gives full consideration to the amplitude, symbol information between center pixel gray value and the local area, mainly includes three kinds of descriptors, i.e., describe the symbol (CLBP_Sign, CLBP_S), the center pixel description operator (CLBP_Center, CLBP_C), the numerical description operator (CLBP_Magnitude, CLBP_M). The definition is expressed as follows,

$$CLBP\_M_{m,r} = \sum_{i=0}^{m-1} s(D_i - T)2^i \tag{8}$$

$$CLBP\_S_{m,r} = \sum_{i=0}^{m-1} s(g_i - g_c)2^i \tag{9}$$

$$CLBP\_C_{m,r} = s(g_c - g_N) \tag{10}$$

where $D_i = g_i - g_c$, $g_N = \frac{1}{N}\sum_{j=0}^{N-1} g_j$, $T = \frac{1}{N}\sum_{j=0}^{N-1}\frac{1}{m}\sum_{i=0}^{m-1}(g_i - g_c)$, and N is the number of sub-windows for image partition. Here, CLBP_S is the same as the traditional LBP definition. CLBP_M compares the difference between the gray-scales amplitude of two pixels and the global gray-scales and describes the gradient difference information of the local window, which reflects the contrast. CLBP_C effectively captures the global contrast information. The information captured by the three operators is complementary to each other.

One of the defects of CLBP algorithm is that only local features of the single scale are described. A possible solution to extract the texture feature of ship image in multiple resolutions, i.e., MS-CLBP, is considered. The MS-CLBP algorithm is improved on the basis of the typical CLBP. Specifically, multiple images with different scales are obtained by down-sampling the original images using double cubic interpolation. Then, the number m and radius r of the neighborhoods are fixed to extract the CLBP_S and CLBP_M histogram features at different scales, which are concatenated to form an MS-CLBP representation. An example of the implementation of an MS-CLBP operator is depicted in Fig. 4.

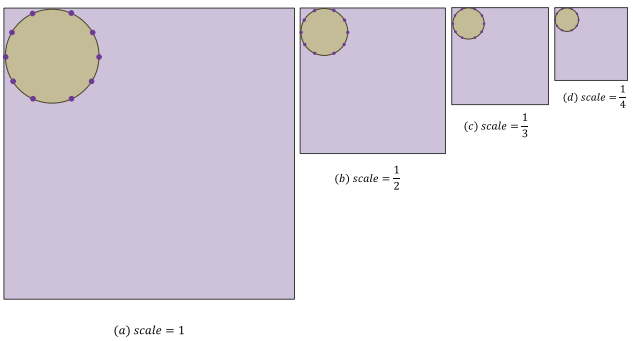

FIGURE 4. An example of a 4-scale CLBP operator(m=10,r=2).

C. FEATURE-LEVEL FUSION

After feature extraction, feature-level fusion strategy is taken consideration to obtain more comprehensive representation, as illustrated in Fig. 5. In the proposed framework, low-level features are extracted by the Gabor-based MS-CLBP which has the ability to capture spatial structure and texture feature. On the other hand, high-level features are extracted by the CNN that can capture the structure and

Algorithm 3 Feature-Level Fusion

Require: Prepared training set and testing set

1: Extract high-level features in Algorithm 1 and low-level features in Algorithm 2

2: Concatenate two types feature into a composite vector

3: Fuse features that are fed into SVM classifier for final classification

Ensure: Experimental results

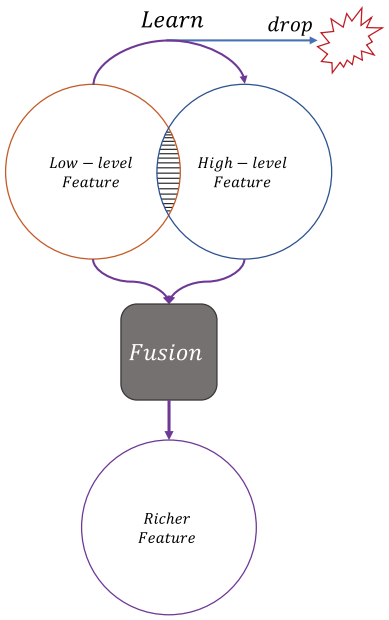

FIGURE 5. The procedure of feature-level fusion strategy.

semantic information. Deep CNN takes advantage of multiple filters to learn global feature via layer-to-layer propagation.

Although the CNN has provided excellent performance in such visual tasks as face recognition and scene classification, it still has the following shortcomings. First of all, deep learning requires a large number of labeled sample to train models and the limitation of the labeled training samples may result in over-fitting. Secondly, the CNN learning feature is based on low-level features to obtain in the first convolution layer which can cause some important information lost, such as edge, contour and so on. Thus, a better low-level feature extractor, such as Gabor filtering and MS-CLBP, can be fused to get more abundant feature representation.

On the one hand, a fine-tuning scheme is taken consideration on the condition of insufficient labeled data; on the other hand, Gabor filter is utilized to extract features in different directions. It is expected that Gabor filter with rotation invariance features or MS-CLBP with multi-scale features can make up the shortcoming of deep CNN. Therefore, it is necessary to employ the feature-level fusion strategy to cascade two types of features into a composite one, producing the more comprehensive representation, i.e., fine-tuning CNN+Gabor+MS-CLBP. The specific description is given in Algo. 3.

Finally, fused features are fed into the typical classifier, i.e., SVM, whose objective function is defined as,

$$\arg\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\max(0, 1 - z_i \mathbf{w}^{\mathrm{T}} x_i) \qquad (11)$$

where the w are the weights and the parameter C is the penalty coefficient that represents the ability to put up with the error. Specifically, the higher the value of C is, the less tolerant of errors and easily resulting in over-fitting are. In contrast, the smaller the C is, the less likely the fitting is. Thus, C is too large or too small and generalization ability is poor. As shown in Fig. 6, we set C=3 for the BCCT200-resize data and C=7 for the VAIS data.

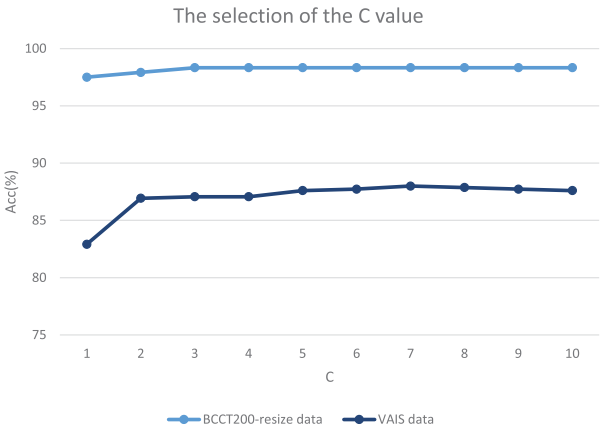

FIGURE 6. The value of the penalty coefficient C for two experimental data.

III. EXPERIMENTAL RESULTS

In order to validate the performance of the proposed method, we conduct several experiments on available ship image data. All the experiments are implemented in Python, MATLAB, and Caffe. The Caffe is a deep learning tool developed by the Berkeley vision and community contributors [52]. Experimental environment is *Ubuntu 14.04*, dual *Intel i5 4590* CPUs, 8GB memory and GPU of *Nvidia GTX 970*.

A. EXPERIMENTAL DATA

The first experimental data is an overhead satellite scene named as BCCT200-resize, which consists of small grey-scale ship images chipped out of larger electro-optical satellite images by RAPIER Ship Detection System as depicted in [24]. Moreover, they have been preprocessed to be rotated and aligned with uniform dimensions and orientation. The dataset includes four categories, i.e., cargos, barges, tankers, and containers. Each ship category has 200 images of size 300×150, as shown in Fig. 7.

The second experimental data is a publicly available data of paired visible and infrared ship images named as VAIS [42]. The dataset consists of 2865 images (1623 visible and 1242 infrared), of which there are 1088 corresponding pairs in total. The dataset includes 6 coarse-grained categories, i.e., merchant ships, sailing ships, mediums

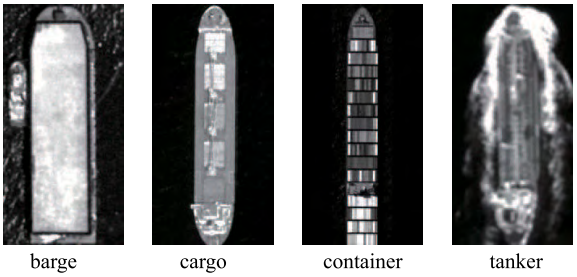

FIGURE 7. Illustration of the BCCT200-resize data.

passenger ships, mediums ''other'' ships, tug boats, and small boats. The area of the visible bounding boxes ranges from 644 to 6350890 pixels, with a mean of 181319 pixels and a median of 13064 pixels, as shown in Fig. 8. The dataset is partitioned into ''official'' train and test groups, resulting in 539 image pairs and 334 singletons for training, and 549 image pairs and 358 singletons for testing. We only choose the visible ship imagery category. In order to facilitate a fair comparison, pre-process of images for deep CNN is implemented the same as [42], and we resize each chip to the size 256×256 by using bicubic interpolation.

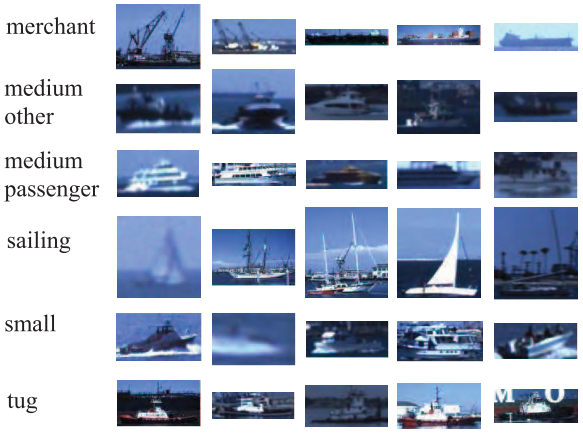

FIGURE 8. Illustration of the VAIS data.

B. PARAMETER TUNING

To train the deep network and avoid over-fitting caused by the limited size of the training dataset, the parameters (weights and biases) of convolution layers and fully-connected layers are transferred from CaffeNet. Then, fine-tuning with a smaller learning rate is accomplished, which avoids training the network from scratch. As for fine-tuning, the learning rate of CNN model plays an important role in the convergence of speed, which can affect the training performance. Specifically, for the BCCT200-resize data, the learning rate set as 0.0001 with the policy of stochastic gradient descent (SGD) [53]. The momentum is 0.9, gamma is 0.9, weight decay is 0.005 and the max iteration is 20000. As for the VAIS data, the learning rate is set as 0.0001 with the policy of SGD [53]. The momentum is 0.95, gamma is 0.9, weight decay is 0.05, and the max iteration is 20000.

The fine-tuning strategy can effectively alleviate the over-fitting problem when the amount of labeled data is not enough. The results of with and without fine-tuning is listed in Table 4. From the results, it is easily to observe that the overall accuracy on the VAIS data has only reached 81.9% without fine-tuning, which is much lower than that of with fine-tuning.

Figs. 9-10 illustrate the Gabor and MS-CLBP features using two experimental data, respectively. According to previous introduction, the spatial frequency bandwidth of Gabor filtering set as 5 and the orientations θ set as $[0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}]$ for both two experimental datasets. The CLBP operator has two very important parameters m (the number of neighbors) and r (the radius), both of them directly affect the final classification accuracy. In our experiments, we tune parameters based on the available training data, and report the results in Fig. 11. Here, we choose $(m, r) = (10, 6)$ for the BCCT200-resize dataset, and $(m, r) = (10, 5)$ for the VAIS dataset. What's more, the scale is set 1/ [1:6]. And the dimensionality of CLBP features concatenated by CLBP_S and CLBP_M histograms for both two datasets is set to 216. Specifically, for each input image, we first obtain the Gabor feature in eight orientations, then MS-CLBP descriptor is utilized to extract local feature based on the Gabor feature in each direction. Finally, the eight MS-CLBP features are cascaded to get the final feature.

C. CLASSIFICATION PERFORMANCE

To validate the performance of the proposed method, we compare with some state-of-the-art methods under the same experimental conditions. Tables 1-2 list available training data and testing data for two experimental data, respectively. The comparison results are depicted in Tables 3-4, respectively. It is obvious that the proposed method achieves more superior classification performance compared with other

TABLE 1. Selected classes for evaluation and the numbers of training and test samples using the BCCT 200-resize data.

No.	Class	Train	Test
1	Barge	140	60
2	Cargo	140	60
3	Container	140	60
4	Tanker	140	60
Total		640	240

TABLE 2. Selected classes for evaluation and the numbers of training and test samples using the VAIS data.

No.	Class	Train	Test
1	Merchant	103	71
2	Medium-other	99	86
3	Medium-Passenger	78	62
4	Sailing	214	198
5	Small	342	313
6	Tug	37	20
Total		873	750

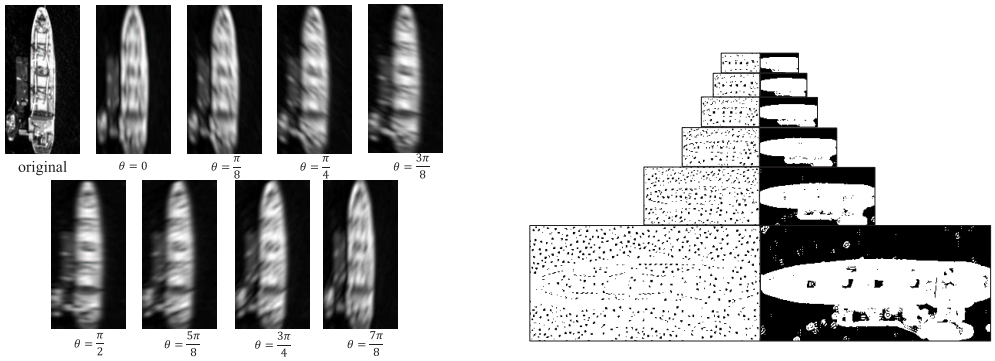

FIGURE 9. Illustration of Gabor and MS-CLBP features using the BCCT200-resize data. (a) Gabor feature with eight orientations. (b) MS-CLBP feature.

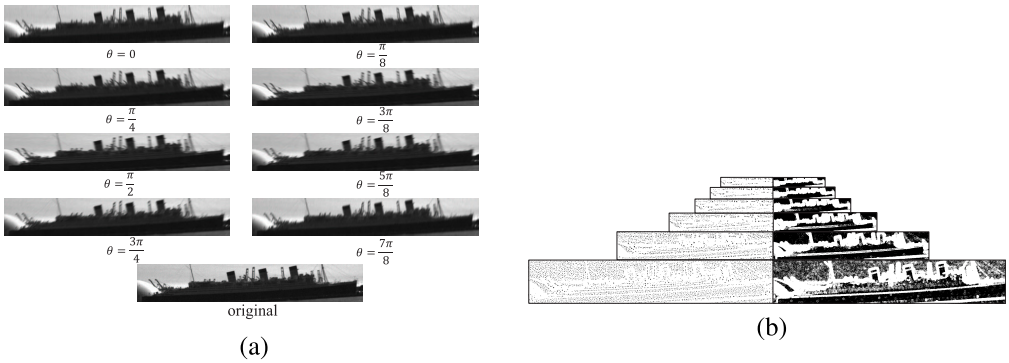

FIGURE 10. Illustration of Gabor and MS-CLBP features using the VAIS data. (a) Gabor feature with eight orientations. (b) MS-CLBP feature.

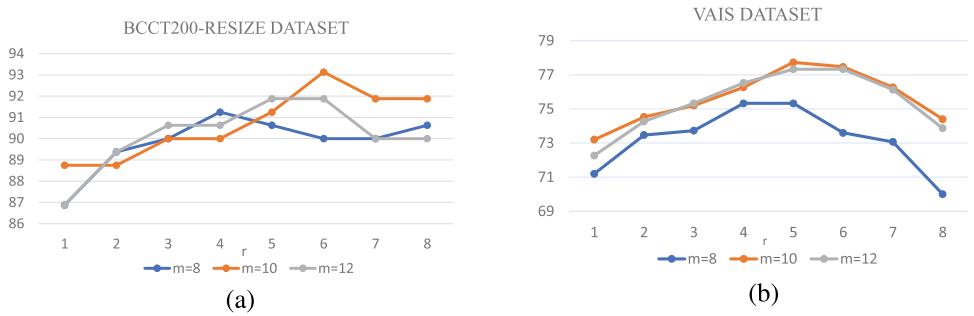

FIGURE 11. Classification accuracy (%) with varying Parameters (m, r) of CLBP for two experimental data. (a) BCCT200-resize data. (b) VAIS data.

existing algorithms. Especially, for the BCCT200-resize data, improvement of the proposed method over the state-of-the-art MFL [22] is 4%. Moreover, the proposed approach gains about 2.5% higher overall accuracy for the VAIS data than the MFL, which verifies the advantages of the fused features. What's more, the sole fine-tuning CNN has better performance when compared to other existing methods because the high-level semantic features are more advanced and discriminative.

In order to verify the enhanced discriminative power of the feature-fusion framework, we further compare the proposed approach with these using individual features. The class-specific classification accuracy is listed in Tables 5-6. It is obvious that the proposed method outperforms all the approaches based individual features. For the BCCT200-resize data set, the low-level feature representation method, i.e., Gabor+MS-CLBP, achieves the highest accuracy for *barge* category, while the high-level feature representation method, i.e., fine-tuning CNN, achieves the highest accuracy for *container* category. The reason is that the inter-class difference among *cargo*, *container* and *tanker* is relatively small, which leads to the difficulty of distinguishing for

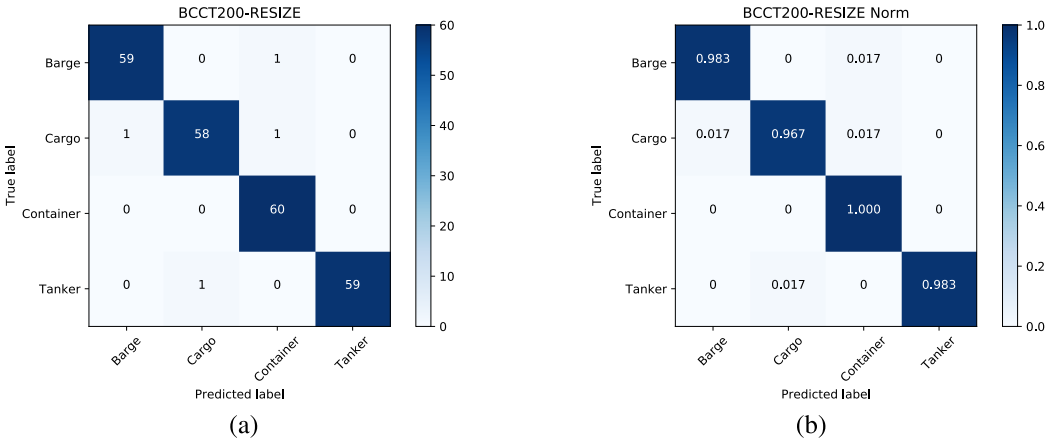

FIGURE 12. Classification confusion matrix of the proposed method using the BCCT200-resize data. (a) Confusion Matrix. (b) Confusion Matrix Normalization.

TABLE 3. Comparison of classification accuracy (%) with some-of-the-art methods for the BCCT200-resize data.

Method	Accuracy(%)
MPCA+SVM [2]	79.1
BOW+SVM [2]	76.8
HOG+SVM [2]	81.6
CNN [37]	94.5
Hierarchical multiscale LBP(HMLBP) [2]	90.8
Gabor+CNN [54]	93.75
Gabor+MS-CLBP+SVM [22]	90.63
MFL(feature-level)+SVM [22]	94.00
Fine-tuning CNN [37]	96.25
Fine-tuning CNN+SVM	97.50
Proposed Method	**98.33**

TABLE 4. Comparison of classification accuracy (%) with some-of-the-art methods for the VAIS data.

Method	Accuracy(%)
Gnostic Field [42]	82.4
HOG+SVM [2]	71.87
CNN [42]	81.9
Gnostic Field+CNN [42]	81.0
Gabor+CNN [54]	80.83
Gabor+MS-CLBP [22]	77.73
MFL(feature-level)+SVM [22]	85.33
Fine-tuning CNN [37]	85.75
Fine-tuning CNN+SVM	86.53
Proposed Method	**88.00**

low-level feature obtained by traditional extraction method. Comparatively, low-level feature performs best to *barge*. Furthermore, the profile of *barge* is more clear than *container*, while *container* class includes more detailed information. MS-CLBP is a better descriptor that obtains local feature, e.g., edge information, while deep CNN can automatically capture the structure information by convolution operation. Therefore, high-level feature obtains better performance for *container* class.

For the VAIS data, it is easily found that *medium-other* and *tug* have low accuracy among all the methods.

TABLE 5. Class-specific accuracy (%) for the BCCT200-resize data.

Class	Gabor+MS-CLBP+SVM	CNN+SVM	Proposed Method
Barge	98.33	96.67	**98.33**
Cargo	78.33	96.67	**96.67**
Container	88.33	100.00	**100.0**
Tanker	93.33	96.67	**98.33**
Execution time(s)	210.07	13	221.07

TABLE 6. Class-specific accuracy (%) and execution time (in Seconds) for the VAIS data.

Class	Gabor+MS-CLBP+SVM	CNN+SVM	Proposed Method
Merchant	67.61	91.55	**92.96**
Medium-other	24.41	50.00	**51.16**
Medium-Passenger	61.29	87.09	**87.09**
Sailing	92.93	94.44	**97.98**
Small	91.69	92.01	**92.65**
Tug	25.00	60.00	**60.00**
Execution time(s)	762.73	77	836.73

Compared with other categories, the number of *Tug* is relatively small, which affects the classification performance. For *medium-Passenger*, consisting of *fishing*, *medium other*, 31.40% of them are mistaken as *small* boats, which includes *speedboat*, *jetski*, *smaller pleasure*, *larger pleasure*, and *small*. As shown in Fig. 8, the quality of images is not good, which is another important factor for performance of *medium-other* ships. Furthermore, some *small* and *medium-passenger* ships have relatively high similarity.

Fig. 12 depicts the confusion matrix of the proposed approach for the BCCT200-resize dataset. As displayed in Fig. 12(a), the diagonal elements of the confusion matrix denote the correct quantity of each category. The class-specific accuracy (%) is shown in Fig. 12(b), where the major confusion occurs between class 2 (i.e., *cargo*) and class 3 (i.e., *container*). We observe that some images of *barge* are similar to these of *container*. Fig. 13 shows the confusion matrix of feature-level fusion for the VAIS dataset. It is apparent that the major confusion occurs within class 2 (i.e., *medium-other*), class 5 (i.e., *small*), class 3

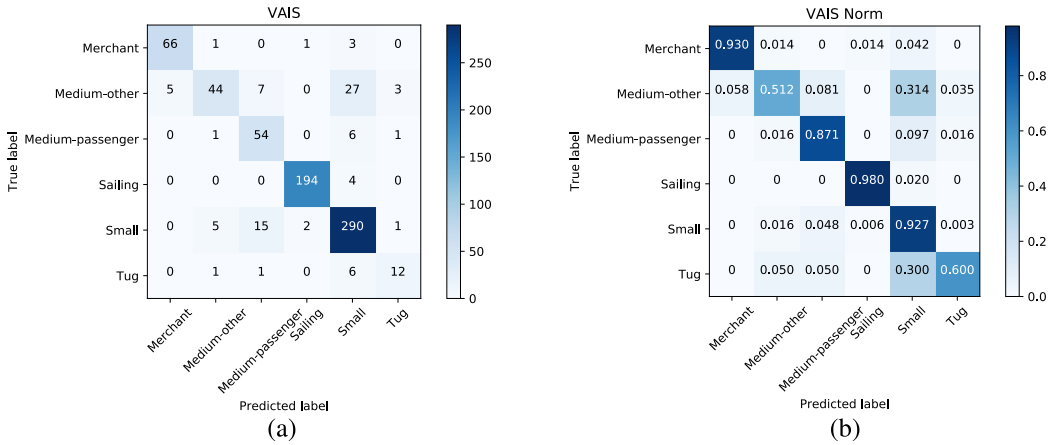

FIGURE 13. Classification confusion matrix of the proposed method using the VAIS data. (a) Confusion Matrix. (b) Confusion Matrix Normalization.

TABLE 7. Classification accuracy (%) in different numbers of training samples for the BCCT200-resize data.

Training set	Test set	Gabor+MS-CLBP+SVM	Fine-tuning CNN	Proposed Method
560(140/per)	240(60/per)	89.58	96.25	**98.33**
480(120/per)	320(80/per)	87.81	95.91	**96.25**
400(100/per)	400(100/per)	87.25	95.00	**95.25**
320(80/per)	480(120/per)	85.42	93.95	**94.58**
240(60/per)	560(140/per)	81.96	93.75	**94.11**
160(40/per)	640(160/per)	78.75	93.12	**93.28**

(i.e., *medium-passenger*), or between class 6 (i.e., *tug*) and class 5 (i.e., *small*). Category of *medium-other* consists of *fishing*, *medium other*, *small boats* includes *speedboat*, *jetski*, *smaller pleasure*, *larger pleasure*, *small*. Therefore, there is high similarity between these ships with small sizes and classes of *medium-other*.

We also conduct an experiment when the amount of training data per class is varied. The sensitivity of the proposed method is investigated over a range of training-data set sizes. In practical situations, the number of training samples available is often insufficient. Take the BCCT20-RESIZE data for example, Table 7 lists the overall accuracy as a function of varying training sample sizes. Under the same condition of training and testing samples, the performance of the proposed method is always better than that of traditional fine-tuning CNN or Gabor + MS-CLBP. Particularly, when the size of training is extremely small (e.g., 40), the proposed approach still achieves the accuracy as high as 93.28%, which confirms its effectiveness.

Then, we conduct statistical analysis of counting numbers of the ship images for each class that are predicted wrongly. As listed in Tables 8-9, Y represents the correctly-classified examples, and N indicates the wrongly-classified examples. The last line of each column represents the sum of ship images satisfying three conditions; for example, the column of (N, N, Y) means that the number of ship images is neither classified by using the features of fine-tuning CNN nor by using the features of Gabor+MS-CLBP but correctly classified by using the features of the proposed method.

TABLE 8. Statistics the number of the vessel image of each types class that are predicted wrongly under different features for the BCCT 200-resize data.

BCCT200-RESIZE data								
Gabor+MS-CLBP	N	N	Y	Y	Y	N	Y	N
Fine-tuning CNN	N	Y	N	Y	Y	Y	N	N
Proposed Method	Y	Y	Y	Y	N	N	N	N
Total	2	19	2	213	0	2	0	2

For the BCCT200-RESIZE data, the proposed approach can correctly classify 2 images that are misclassified by using two other features, and for the VAIS data, the number is 7.

The statistical significance about the performance improvement of the proposed method can be evaluated by using the standardized McNemar's test. When the Z value of McNemar's test is larger than 1.96 and 2.58, it means that the two results are statistically different in the confidence level of 95% and 99%, respectively. The sign of Z denotes that whether the first classifier outperforms the second classifier ($Z>0$). In our experiments, the comparison between proposed approach and other existing methods is made separately. As listed in Table 10, all values are larger than 2.58, which demonstrates the effectiveness of the proposed method.

We further enumerate the classification results under different types of features as illustrated in Fig. 14. Take the BCCT200-RESIZE data for example, the red solid box represents the correct image to be predicted, and the red dashed box indicates the images that are misclassified. When the

TABLE 9. Statistics the number of the vessel image of each types class that are predicted wrongly under different features for the VAIS data.

VAIS data								
Gabor+MS-CLBP	N	N	Y	Y	Y	N	Y	N
Fine-tuning CNN	N	Y	N	Y	Y	Y	N	N
Proposed Method	Y	Y	Y	Y	N	N	N	N
Total	7	94	8	551	1	3	23	63

TABLE 10. Statistical significant evaluated by the McNemar's test based on difference between methods.

BCCT200-resize	VAIS
mean Z/significant?	mean Z/significant?
Proposed Method vs Gabor+MS-CLBP	
12.96/yes	13.96/yes
Proposed Method vs Fine-tuning CNN	
13.30/yes	13.80/yes
Proposed Method vs Gabor+CNN	
11.53/yes	13.60/yes

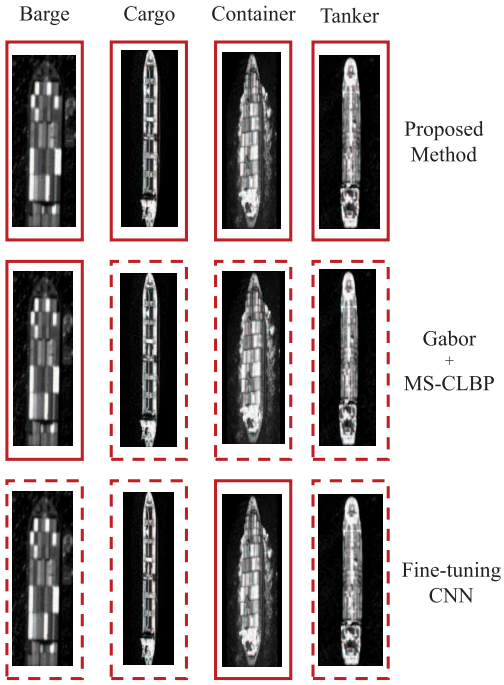

FIGURE 14. Illustration of each class classification performance for the BCCT200-RESIZE data.

result of fine-tuning CNN or Gabor+MS-CLBP is erroneous (e.g., *Cargo* and *tanker*), the proposed method can predict the category of ship correctly, which further confirms its effectiveness.

IV. CONCLUSIONS

In this paper, an effective ship classification framework was proposed by learning the high-level features via deep CNN and incorporating the multi-scales rotation invariance features. Traditional CNN has some drawbacks (e.g., ignoring part of local features and sufficient labeled data for training), which result in that the direct application in ship classification tasks is difficult. In the proposed strategy, Gabor filter was employed to obtain features in different directions, and MS-CLBP was applied to obtain the local texture, spatial and profile information of ship images, which compensated the shortcoming of deep CNN. Furthermore, the fused level representation has been demonstrated more powerful and comprehensive due to both high-level and low-level features are complementary to each other. Through experimental analysis, the proposed approach achieved the average accuracy of 98.33% on the BCCT200-RESIZE data and 88.00% on the VAIS data, which has been confirmed to be superior to state-of-the-art methods.

REFERENCES

[1] C. M. Pilcher and A. Khotanzad, "Maritime ATR using classifier combination and high resolution range profiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 4, pp. 2558–2573, Oct. 2011.
[2] K. Rainey, S. Parameswaran, and J. Harguess, "Vessel classification in overhead satellite imagery using learned dictionaries," in *Proc. Int. Soc. Opt. Photon. Appl. Digit. Image Process. XXXV*, 2012, p. 84992F.
[3] N. Rowell, S. Parkes, and M. Dunstan, "Image processing for near earth object optical guidance systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 1057–1072, Apr. 2013.
[4] O. Kechagias-Stamatis, N. Aouf, and M. A. Richardson, "3D automatic target recognition for future LIDAR missiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2662–2675, Dec. 2016.
[5] J. Zou, W. Li, C. Chen, and Q. Du, "Scene classification using local and global features with collaborative representation fusion," *Inf. Sci.*, vol. 348, pp. 209–226, Jun. 2016.
[6] C. Zhang, G. Zhu, Q. Huang, and Q. Tian, "Image classification by search with explicitly and implicitly semantic representations," *Inf. Sci.*, vol. 376, pp. 125–135, Jan. 2017.
[7] G. K. Yüksel, B. Yalıtuna, Ö. F. Tartar, F. Adlı, K. Eker, and O. Yörük, "Ship recognition and classification using silhouettes extracted from optical images," in *Proc. Signal Process. Commun. Appl. Conf.*, 2016, pp. 1617–1620.
[8] X. Zhang, H. Wang, and W. Cheng, "Vessel detection and classification fusing radar and vision data," in *Proc. 7th Int. Conf. Inf. Sci. Technol. (ICIST)*, 2017, pp. 474–479.
[9] F. Teng and Q. Liu, "Robust multi-scale ship tracking via multiple compressed features fusion," *Signal Process., Image Commun.*, vol. 31, pp. 76–85, Feb. 2015.
[10] W. Guo, X. Xia, and X. Wang, "A remote sensing ship recognition method based on dynamic probability generative model," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6446–6458, 2014.
[11] Q. Oliveau and H. Sahbi, "Learning attribute representations for remote sensing ship category classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2830–2840, Jun. 2017.
[12] V. F. Arguedas, "Texture-based vessel classifier for electro-optical satellite imagery," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3866–3870.
[13] S. Park, C. J. Cho, B. Ku, S. Lee, and H. Ko, "Simulation and ship detection using surface radial current observing compact HF radar," *IEEE J. Ocean. Eng.*, vol. 42, no. 3, pp. 544–555, Jul. 2017.
[14] J. Barnum, "Ship detection with high-resolution HF skywave radar," *IEEE J. Ocean. Eng.*, vol. JOE-11, no. 2, pp. 196–209, Apr. 1986.
[15] W. Guo, X. Xia, and X. Wang, "Variational approximate inferential probability generative model for ship recognition using remote sensing data," *Optik-Int. J. Light Electron Opt.*, vol. 126, no. 23, pp. 4004–4013, 2015.
[16] L. Chunyan, Z. Huanxin, S. Hao, and Z. Shilin, "Combing rough set and RBF neural network for large-scale ship recognition in optical satellite images," in *Proc. IOP Conf. Ser., Earth Environ. Sci.*, vol. 17, no. 1, 2014, p. 012218.
[17] W. Guo, X. Xia, and X. Wang, "A remote sensing ship recognition method of entropy-based hierarchical discriminant regression," *Optik*, vol. 126, no. 20, pp. 2300–2307, Oct. 2015.

[18] L. Huang, C. Chen, W. Li, and Q. Du, "Remote sensing image scene classification using multi-scale completed local binary patterns and Fisher vectors," *Remote Sens.*, vol. 8, no. 6, p. 483, 2016.
[19] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.
[20] Z. Guo, L. Zhang, D. Zhang, and X. Mou, "Hierarchical multiscale LBP for face and palmprint recognition," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 4521–4524.
[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 886–893.
[22] L. Huang, W. Li, C. Chen, F. Zhang, and H. Lang, "Multiple features learning for ship classification in optical imagery," *Multimedia Tools Appl.*, vol. 77, no. 11, pp. 13363–13389, 2018.
[23] J. Harguess and K. Rainey, "Are face recognition methods useful for classifying ships?" in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, Oct. 2011, pp. 1–7.
[24] K. Rainey and J. Stastny, "Object recognition in ocean imagery using feature selection and compressive sensing," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, Oct. 2011, pp. 1–6.
[25] C.-Y. Low, A. B.-J. Teoh, and C.-J. Ng, "Multi-fold Gabor filter convolution descriptor for face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 2094–2098.
[26] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.
[27] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.
[28] J. Li, T. Wang, Y. Zhou, Z. Wang, and H. Snoussi, "Using Gabor filter in 3D convolutional neural networks for human action recognition," in *Proc. 36th Chin. Control Conf.*, 2017, pp. 11139–11144.
[29] Y. Hu, C. Li, D. Hu, and W. Yu, "Gabor feature based convolutional neural network for object recognition in natural scene," in *Proc. Int. Conf. Inf. Sci. Control Eng.*, 2016, pp. 386–390.
[30] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[31] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 770–778.
[33] Z. Chen, C. Ying, C. Lin, S. Liu, and W. Li, "Multi-view vehicle type recognition with feedback-enhancement multi-branch CNNs," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[34] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral–spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.
[35] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 2449–2457.
[36] G. Cheng, P. Zhou, and J. Han, "RIFD-CNN: Rotation-invariant and Fisher discriminative convolutional neural networks for object detection," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 2884–2893.
[37] K. Rainey, J. D. Reeder, and A. G. Corelli, "Convolution neural networks for ship type recognition," in *Proc. Int. Soc. Opt. Photon. Autom. Target Recognit. XXVI*, 2016, p. 984409.
[38] J. Wang, C. Luo, H. Huang, H. Zhao, and S. Wang, "Transferring pre-trained deep CNNs for remote scene classification with general features learned from linear PCA network," *Remote Sens.*, vol. 9, no. 3, p. 225, 2017.
[39] G.-S. Xie, X.-Y. Zhang, S. Yan, and C.-L. Liu, "Hybrid CNN and dictionary-based models for scene recognition and domain adaptation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1263–1274, Jun. 2017.
[40] J. Lu, G. Wang, and P. Moulin, "Localized multifeature metric learning for image-set-based face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 529–540, Mar. 2016.
[41] L. Cui, F. Meng, Y. Shi, M. Li, and A. Liu, "A hierarchy method based on LDA and SVM for news classification," in *Proc. IEEE Int. Conf. Data Mining Workshop*, Dec. 2015, pp. 60–64.
[42] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, and C. Kanan, "VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 10–16.
[43] D.-X. Xue, R. Zhang, H. Feng, and Y.-L. Wang, "CNN-SVM for microvascular morphological type recognition with data augmentation," *J. Med. Biol. Eng.*, vol. 36, no. 6, pp. 755–764, 2016.
[44] K. Mori, M. Matsugu, and T. Suzuki, "Face recognition using SVM fed with intermediate output of CNN for face detection," in *Proc. IAPR Conf. Mach. Vis. Appl.*, 2009, pp. 410–413.
[45] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, 2012.
[46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.
[47] C. Chen, L. Zhou, J. Guo, W. Li, H. Su, and F. Guo, "Gabor-filtering-based completed local binary patterns for land-use scene classification," in *Proc. IEEE Int. Conf. Multimedia Big Data (BigMM)*, Apr. 2015, pp. 324–329.
[48] P. V. Bankar and A. C. Pise, "Face recognition by using Gabor and LBP," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, 2015, pp. 0045–0048.
[49] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological*, vol. 61, no. 2, pp. 103–113, 1989.
[50] Y. Zhong and H. Li, "Is block matching an alternative tool to LBP for face recognition?" in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2015, pp. 723–727.
[51] A. Porebski, N. Vandenbroucke, and D. Hamad, "LBP histogram selection for supervised color texture classification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2014, pp. 3239–3243.
[52] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
[53] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*. New York, NY, USA: Springer, 2012.
[54] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.

QIAOQIAO SHI received the B. S. degree from the Beijing University of Chemical Technology, Beijing, China, in 2016, where she is currently pursuing the M. S. degree, under the supervision of Dr. W. Li.

WEI LI received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Post-Doctoral Researcher with the University of California, Davis, CA, USA. He is currently with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.

Dr. Li received the 2015 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society for his service for IEEE JSTARS. He is an Active Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He serves as the Guest Editor for special issue of the *Journal of Real-Time Image Processing*, *Remote Sensing*, and the IEEE JSTARS.

FAN ZHANG received the B.E. degree in communication engineering from the Civil Aviation University of China, Tianjin, China, in 2002, the M.S. degree in signal and information processing from Beihang University, Beijing, China, in 2005, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Science, Beijing, in 2008.

He is currently an Associate Professor of electronic and information engineering with the Beijing University of Chemical Technology, Beijing. His research interests are synthetic aperture radar signal processing, high performance computing, and machine learning.

Dr. Zhang has been a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the *International Journal of Antennas and Propagation*.

WEI HU received the B.S. and M.S. degrees in computer science from the Dalian University of Science and Technology, Dalian, China, in 1999 and 2002, respectively, and the Ph.D. degree in computer science from Tsinghua University, Beijing, China, in 2006.

He is currently an Associate Professor of computer science with the Beijing University of Chemical Technology, Beijing. His research interests are computer graphics, computational photography, and scientific visualization.

Dr. Hu has been a Reviewer for the IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, *EuroGraphics*, and *Pacific Graphics*.

XU SUN is currently an Associate Professor with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, CAS. His research interests include hyperspectral image processing and information extraction.

LIANRU GAO received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He is currently a Professor with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, CAS. His research interests include hyperspectral image processing and information extraction. In last ten years, he led 10 scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China, and by the Key Research Program of the CAS and so on. He has published over 110 peer-reviewed papers, including 47 journal papers. He was co-authored an academic book *Hyperspectral Image Classification And Target Detection*. He obtained 12 National Invention Patents and four Software Copyright Registrations. His core research achievements gained the Outstanding Science and Technology Achievement Prize of the CAS in 2016. He received the China National Science Fund for Excellent Young Scholars in 2017. He received the recognition of the Best Reviewers of the IEEE JSTARS in 2015.