An Improved InceptionV3 Network for Obscured Ship Classification in Remote Sensing Images

Kun Liu, Shengtao Yu, and Sidong Liu

***Abstract*—Ship target classification plays an important role in tasks such as maritime traffic control, maritime target tracking, and military reconnaissance. The complex ocean environment often causes obscuration of the ship targets, thus resulting in low accuracy of the obscured targets. This article presents a novel target classification algorithm—improved InceptionV3 and center loss convolution neural network (IICL-CNN)—based on the well-established inception network to improve the accuracy of obscured targets. This algorithm features a new objective function, which is designed to learn common features of both the clear samples and the obscured samples and, in the meantime, reduce the intraclass distance among the obscured samples. Experiments were performed on an optical remote sensing image dataset which consisted of 48 000 ship images in nine categories. The proposed method demonstrated superior performance on the obscured ship targets compared to the original InceptionV3 model. On average, the accuracy was 4.23%, 5.98%, and 17.48% higher on the ship targets that were occluded by levels of 30%, 50%, and 70%, respectively. Our experimental results showed that the proposed IICL-CNN could effectively improve the accuracy of the ship targets at various occlusion levels.**

***Index Terms*—Convolution neural network (CNN), deep learning, feature extraction, fog occlusion, image classification, remote sensing image.**

I. Introduction

SHIP target classification is an emerging research area which has a broad range of applications in civil and military fields, such as maritime traffic control, onshore search, maritime target tracking, and military reconnaissance. With the developments in computer vision and machine learning, there are many recent studies in this area. Zhu *et al*. [1] recently proposed a method for ship identification, which extracted multiple high-dimensional local features from ship images and then used support vctor machine (SVM) for target classification. In order to identify ship targets more accurately, SVM is combined with other methods to improve effectiveness and robustness [2]–[5]. Aiming at the sparseness of targets in optical remote sensing images, Li *et al*. [6] proposed a method for target detection and classification based on morphological matching and machine learning. Gao *et al*. [7] proposed a ship classification algorithm based on hierarchical salient regions. Yu *et al*. [8] proposed an algorithm for uncertain ship target extraction based on a multifeature dynamic fusion model and variance features of optical remote sensing images. Dan *et al*. [9] proposed a transduction learning framework to solve the problem of ship target classification in infrared images. Sun *et al*. [10] proposed an automatic ship target classification method based on scale invariant feature transform (SIFT). The methods proposed in these studies were mainly based on hand-crafted features, such as the position of the main structure of the vessel, length and area of the vessel, space moment, etc. However, these methods could only capture high-level features of the images, possibly ignoring pixel-level details that are also important for target classification.

Recent advances in convolutional neural network (CNN) have reshaped the research in this area and suggested an alternative solution. CNN is capable of automatically learning features of images end-to-end and therefore has developed rapidly in the field of target classification. In 2012, Alex Krizhevsky proposed the CNN model–AlexNet [11] and won the ILSVRC visual field competition with a big lead over the traditional methods. Gong *et al*. [12] refined the AlexNet network to achieve rotation invariance of the target. In 2014, GoogLeNet [13] was proposed and won the ILSVRC championship, with a top-5 error rate of 6.7%. Szegedy *et al*. [14] indicates that very deep convolutional networks have become the mainstream of computer vision since 2014, and have achieved many improvements in various benchmark tests. The increasing volume of labeled data, complemented with advances in model structure, can quickly improve model quality in most tasks. The authors actively explore ways to extend the network, aiming to make use of the added calculations as efficiently as possible through proper decompression and active regularization. Using a model with 5 billion multiply-add operations per inference process and less than 2.5 million total parameters, the method in [14] achieved single-frame evaluations with 21.2% top-1 and 5.6% top-5 error rates. Since then, the inception structure of GoogLeNet network has experienced a few major updates [14], [15] with many variants, and is still under active development. With its good stability and robustness, the target classification methods based on the different inception versions are becoming more popular [16]–[19].

Due to the diversity of ship size and rotation, ship detection from optical remote sensing images is still a difficult task. In order to solve this problem, some of the studies have

Manuscript received April 22, 2020; revised June 29, 2020 and July 29, 2020; accepted August 8, 2020. Date of publication August 18, 2020; date of current version August 28, 2020. This work was supported in part by the Aeronautical Science Foundation of China under Grant 201955015001. *(Corresponding authors: Shengtao Yu; Sidong Liu.)*

Kun Liu and Shengtao Yu are with the School of Information Engineering, Shanghai Maritime University, Shanghai 201306, China (e-mail: kunliu@shmtu.edu.cn; 1095962982@qq.com).

Sidong Liu is with the Macquarie University, Sydney, NSW 2109, Australia (e-mail: sidong.liu@mq.edu.au).

Digital Object Identifier 10.1109/JSTARS.2020.3017676

implemented the effectiveness of ship detection by adjusting the CNN structure [20]–[22]. Lei *et al*. [23] proposed a post-convolution neural network method to extract ships from high-resolution optical remote sensing images. Zhang *et al*. [24] proposed a CNN- extreme learning machine (ELM) network structure combining a convolutional neural network and an ELM. This is used for ship high-resolution distance profile (HRRP) target classification and automatic extraction of deep features from data. Zou *et al*. [25] developed a CNN model with singular value decomposition (SVD) to improve feature extraction for offshore vessels. Chan *et al*. [26] proposed the PCANet model based on principal component analysis (PCA), which learned the multilayer kernels, binary hash codes, and block histograms for image down sampling and feature encoding. Cheng *et al*. [27] proposed a geometric feature extraction method based on iterative linear regression to improve the parameter estimation of a ship target slice. Zhao *et al*. [28] proposed a network architecture based on faster R-CNN to improve the detection performance of the model by using compression and excitation mechanisms. Kang *et al*. [29] proposed a multilayer fusion convolutional neural network based on context region for synthetic aperture radar (SAR) ship detection. But the performances of these CNN-based models are usually compromised by the ship targets in images obscured by the complex meteorological environment on the sea surface. For example, fog occlusion is one of the common situations encountered in the sea environment. Ship targets obscured by the fog are difficult to recognize and therefore can lead to low accuracy or misjudgments. One indirect solution to alleviate the impact of fog occlusion is to remove the fog component from images before feeding them to a CNN model. For example, Yoon *et al*. [30] presented a weighted adaptive image defogging method by extracting features in the RGB channels. Yao *et al*. [31] proposed a method to calculate the defogging rate from the effective edge strength, proving its effectiveness and robustness. Park *et al*. proposed a novel depth-based stereo image dehazing method [32]. The study [33] proposes improved methods of atmospheric light estimation and transmission for the shortcomings of inaccurate atmospheric light estimation. Yoon *et al*. [34] proposed a dehazing image enhancement algorithm for drone images based on a wavelength adaptive image formation model and geometric classification. However, we believe that this challenge can be directly addressed by refining the CNN model to jointly learn the features from both fog occluded ship targets and the clear ship targets.

Therefore, in this study we propose a novel target classification method based on a refined InceptionV3 network to improve the accuracy of the obscured targets. The contribution of this work is threefold. First, a specific dataset of 46 000 ship images with various fog occlusion conditions was established to benchmark different methods. Second, we added a new constraint to the original Softmax cross-entropy function of the CNN to increase the similarity between features learned from the occlusion samples and the clear samples, therefore allowing the features to be shared across the samples with various levels of occlusion. Finally, a center loss function was added to the objective function to reduce the intraclass distances, which further improved the discriminating power of the learned features thereby increasing the accuracy of the obscured samples. The proposed method demonstrated superior performance over previous methods in ship target classification and implies a high applicability in tasks like maritime traffic control, onshore search, maritime target tracking, and military reconnaissance.

II. Fog Occluded Ship Classification Based on Convolution Neural Network

A. Fog Occluded Ship Dataset

A large number of ship images with various fog occlusion conditions are required to enable a CNN model to learn the features of the occluded ship targets. With only a few publicly available ship image datasets, we were motivated to establish our own dataset. The clear ship samples, mostly optical remote sensing images, were obtained using Google crawler, whereas the occluded ship samples in this dataset were obtained by adding Berlin noise [35], [36] to the clear samples to simulate the fog occlusion effect with a weight coefficient controlling the degree of occlusion.

The fog map formation algorithm described in [36] was used to simulate the fog occluded ship images. A simulated fog image is defined as

$$I(x) = D(x)t(x) + A(1 - t(x)) \tag{1}$$

where $I(x)$ is the generated fog image, $D(x)$ is the original image, A is the global atmospheric light component, and $t(x)$ is the medium transmission defined by users to describe the ability of foggy air to transmit light. We first identified the minimum value of RGB channels of the input image to obtain a dark primary channel image, which was then used to estimate the global atmospheric light A through average filtering. To produce the obscured ship samples, a set of K occlusion levels $\varphi = \{\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_k\}$ and occlusion transforms $T_\varphi = \{T_{\varphi_1}, T_{\varphi_2}, T_{\varphi_3}, \ldots, T_{\varphi_k}\}$ were defined, where T_{φ_k} represents a transform with an occlusion level φ_k. Each occlusion transform in T_φ was then applied to all training samples $X = \{x_1, x_2, \ldots, x_N\}$ to derive the obscured samples $T_\varphi X = \{T_{\varphi_1} x_1, T_{\varphi_2} x_2, T_{\varphi_3} x_3, \ldots, T_{\varphi_k} x_N\}$, where a single occlusion sample was defined as $T_\varphi x_i = \{T_{\varphi_k} x_i | k = 1, 2, \ldots, K\}$. The clear samples and their corresponding occluded samples, $\chi = \{X, T_\varphi X\}$, were then used together to train and test the proposed model. In this study, we set $K = 5$ and $\varphi = \{30, 40, 50, 60, 70\}$.

The dataset used contains 46 000 samples across nine types of ships, including aircraft carriers (5196 samples), bulk carriers (5070), cruise ships (5100), warships (5112), submarines (5088), container ships (5130), tankers (5094), barges (5106), and small boats (5106). We split the dataset into 40 000 training images, $\sim$4000 validation images, and $\sim$2000 test images. All images were standardized by rescaling to an image size of 299 $\times$ 299 pixels and applying zero-mean normalization. Fig. 1 illustrates a few clear ships and simulated occluded ships from the dataset.

B. Improved InceptionV3 and Center Loss-CNN

InceptionV3 network can be efficiently decomposed into small convolution kernels, which greatly reduces the number

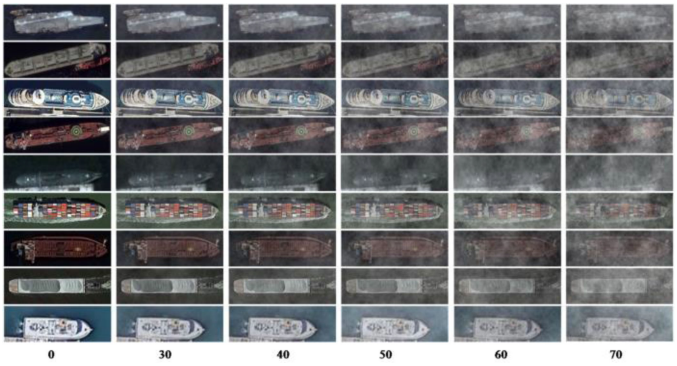

Fig. 1. Examples of the generated fog obscured ship images at different occlusion levels.

of parameters of the model and the chance of overfitting. Meanwhile, it also increases the nonlinear expression ability of the network by extracting features at different abstraction levels and encoding them to high-level features [37]. In this article, we propose two improvements to the InceptionV3 network. First, a fully connected layer (FC layer) was added to the network to merge the features learned from clear targets and occluded targets. Second, a center loss function was added to the loss function to reduce the distance between classes. Therefore, the proposed network is named as improved InceptionV3 and center loss CNN (IICL-CNN).

Fig. 2 shows the structure of the proposed IICL-CNN, where *i* is the *i*th sample, *m* is the category, and *N* is the number of samples. This model consists of three different types of inception modules, with 3, 5, and 3 inception structures, respectively [13]. Although the model has 43 layers, the number of parameters is only one-twelfth of that of AlexNet, which has a great advantage in model deployment. The network configuration is shown in Table I.

IICL-CNN, like other CNN models, can extract the low-level features from the input images using its convolutional and pooling layers. The low-level features can be transformed to high-level abstract features through the inception modules, and further transformed to a linear feature vector using the linear and FC layers. The main function of the FC layer is to merge the features learned from samples with and without fog occlusion and to reduce the impact of feature divergence on target classification.

The process of training IICL-CNN includes calculating the deviation and characteristic error between the predicted value and the ground real situation through forward propagation after the initialization of random parameters. The deviation and the characteristic error are further tested to ensure that they fall within the allowable range, followed by weight update through backpropagation. The network parameters and thresholds are fixed and saved when the termination criteria are met. Adam optimizer [38] is used with a learning rate of 0.001, momentum of 0.9, super parameter β_1 of 0.9, β_2 of 0.999, epsilon ϵ of 10^{-8}, batch size of 32, and 30 000 iterations. To evaluate a trained IICL-CNN model, test samples are taken as inputs to obtain the output predictions, and the average accuracy of the network is calculated by comparing the predictions with the ground truth labels of the test samples.

C. Constrained Objective Function With Center Loss Function

We hypothesize that, if the difference between features learned from the clear samples and the occlusion samples is reduced, or in other words, the features are commonly applicable to, both, clear samples and obscured samples, then the accuracy for the occlusion samples will increase. Therefore, we added a constraint to the objective function to enforce the similarity between clear sample features and occlusion sample features. In addition, since the Softmax cross-entropy loss function can only expand the distance between classes but cannot reduce the intraclass distance, a center loss function was added to control the intraclass distances to further improve the discriminating power of the model. The proposed objective function is defined in (2)

$$L = L_s + \lambda_1 L_l + \lambda_2 L_c \tag{2}$$

where λ_1 and λ_2 are the tradeoff coefficients of the objective function. L_s is a Softmax cross-entropy loss function as defined in (3)

$$L_s = -\frac{1}{N+NK}\sum_{x_i\in\chi}\sum_{k=1}^{C} y_{x_i}[k]\cdot\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{C} e^{W_j^T x_j + b_j}}[k] \tag{3}$$

where y_{x_i} represents the ground truth label of the sample, *N* is the total number of initial training samples in *X*, *K* is the total number of occlusion transformations per $x_i \in X$, and *C* is the total category.

L_l is the objective function feature similarity constraint, which is applied to enforce similarity between features of the clear sample *X* and occlusion samples $T_\varphi X$. In this study, the logarithmic hyperbolic cosine function is chosen as the constraint function as it is a smoother loss function than the L_2 loss function. When the sample characteristic error is small, logcosh(*x*) is similar to the L_2 loss function. When the error is large, logcosh(*x*) is similar to abs(*x*)-log2 and abs(*x*)−log2 is similar to the Huber function when $\delta = 1$. Logarithmic hyperbolic cosine function ensures that the loss function has a continuous derivative. Fig. 3 shows the comparison of the logarithmic hyperbolic cosine function and other functions.

The hyperbolic cosine function uses mean square error (mse) to evaluate the similarity of clear sample features and occlusion sample features. The error is minimized during model training to enforce the model to learn the common features that are shared by both clear and occlusion samples, thereby improving the robustness of the model. L_l is defined as

$$L_l = \frac{1}{N}\sum_{x_i\in X}\log\cosh(\mathrm{O}_{FC}(x_i) - \overline{\mathrm{O}_{FC}(\mathrm{T}_\varphi x_i)}) \tag{4}$$

where *N* is the total number of initial training samples in *X*; $O_{FC}(x_i)$ is the output of the FC layer and also the IICL-CNN feature of the training sample x_i; $\overline{\mathrm{O}_{FC}(\mathrm{T}_\varphi x_i)}$ represents the averaged IICL-CNN features of the occluded training sample

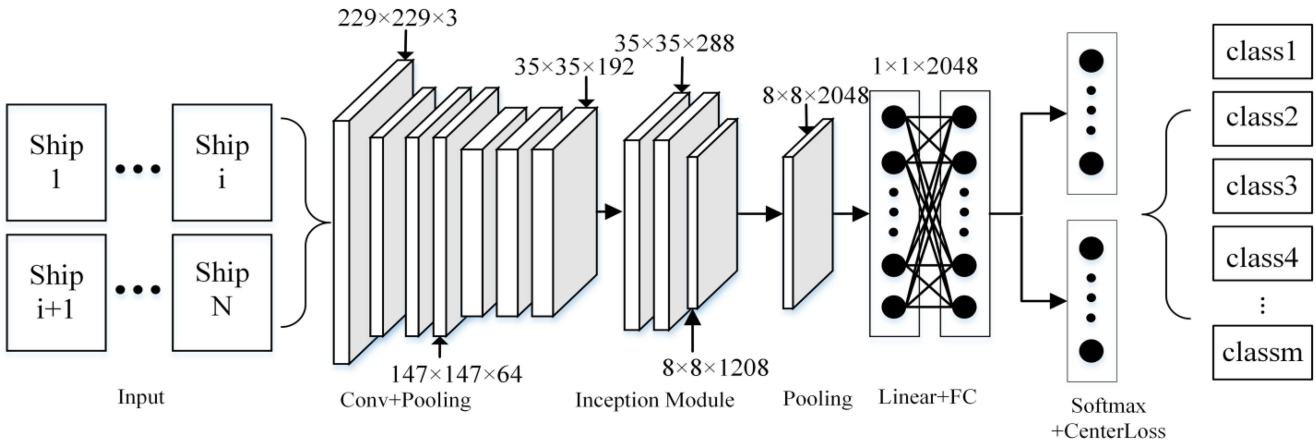

Fig. 2. Ship classification simplified model based on IICL-CNN.

TABLE I
IICL-CNN NETWORK'S CONFIGURATION

Type	Patch size/stride	Input size
conv	3×3/2	299 × 299 × 3
conv	3×3/1	149 × 149 × 32
Conv padded	3×3/1	147 × 147 × 32
pool	3×3/2	147 × 147 × 64
Conv	3×3/1	71 × 71 × 80
Conv	3×3/2	35 × 35 × 192
Conv	3×3/1	35 × 35 × 288
Inception	-	8 × 8 × 1280
pool	88	8 × 8 × 2048
Linear+FC	Logits	1 × 1 × 2048
Softmax+Centerloss	classifier	1 × 1 × m

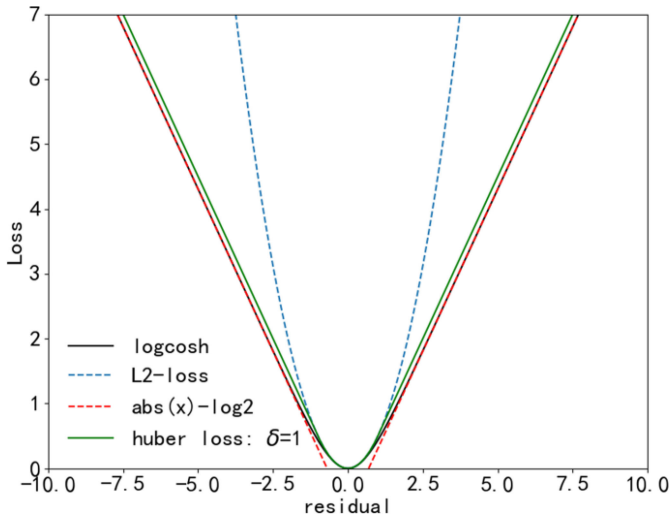

Fig. 3. Comparison of logarithmic hyperbolic cosine functions.

$T_\varphi x_i$

$$\overline{\mathrm{O}_{FC}(\mathrm{T}_\varphi x_i)} = \frac{1}{\mathrm{K}}(\mathrm{O}_{FC}(\mathrm{T}_{\varphi_1} x_i)+\mathrm{O}_{FC}(\mathrm{T}_{\varphi_2} x_i) + \cdots + \mathrm{O}_{FC}(\mathrm{T}_{\varphi_K} x_i)). \tag{5}$$

To reduce the intraclass distance within each class, a center loss function [39] is added to impose the feature discriminant degree of the model, thereby improving the accuracy of the occlusion sample

$$L_c = \frac{1}{2}\sum_{i=1}^{m} ||O_{FC}(\chi_i) - c_{y_i}||_2^2. \tag{6}$$

Among them, $O_{FC}(\chi_i)$ is the feature of the FC layer, and c_{y_i} represents a feature center of the y_ith category.

The class center was derived by averaging the features of the corresponding class. To avoid unnecessary fluctuations caused by a small number of erroneously labeled samples, a scalar was then to control the learning rate [0, 1] when calculating the center loss. The updated calculation of the echelon sum with respect to the characteristics is shown in (7) and (8)

$$\frac{\partial L_c}{\partial O_{FC}(\chi_i)} = O_{FC}(\chi_i) - c_{y_i} \tag{7}$$

$$\Delta c_j = \frac{\sum_{i=1}^{m} \delta(y_i - j)(c_j - O_{FC}(\chi_i))}{1 + \sum_{i=1}^{m} \delta(y_i = j)} \tag{8}$$

where δ(condition) $= 1$ if the condition is satisfied, and δ (condition) $= 0$ if not. Parameters $\{\mathrm{c}_j|\mathrm{j} = 1, 2, \ldots, \mathrm{n}\}$in loss layer. Reorganizing (2) to (6) we can get the objective function

$$L = -\frac{1}{N+NK}\sum_{x_i \in \chi}\sum_{k=1}^{C} y_{x_i}[k] \cdot \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{C} e^{W_j^T x_j + b_j}}[k] + \frac{\lambda_1}{N}\sum_{x_i \in X} \log\cosh(\mathrm{O}_{FC}(x_i) - \overline{\mathrm{O}_{FC}(\mathrm{T}_\varphi x_i)}) + \frac{\lambda_2}{2}\sum_{i=1}^{m} ||O_{FC}(\chi_i) - c_{y_i}||_2^2 \tag{9}$$

III. RESULTS

A. Comparison Between IICL-CNN and InceptionV3

To compare the difference between the features learned by InceptionV3 and IICL-CNN, the features for each sample selected from the test set were extracted at the linear layer of the two networks. Fig. 4 illustrates the extracted high-dimensional features, which were processed using the principal component analysis

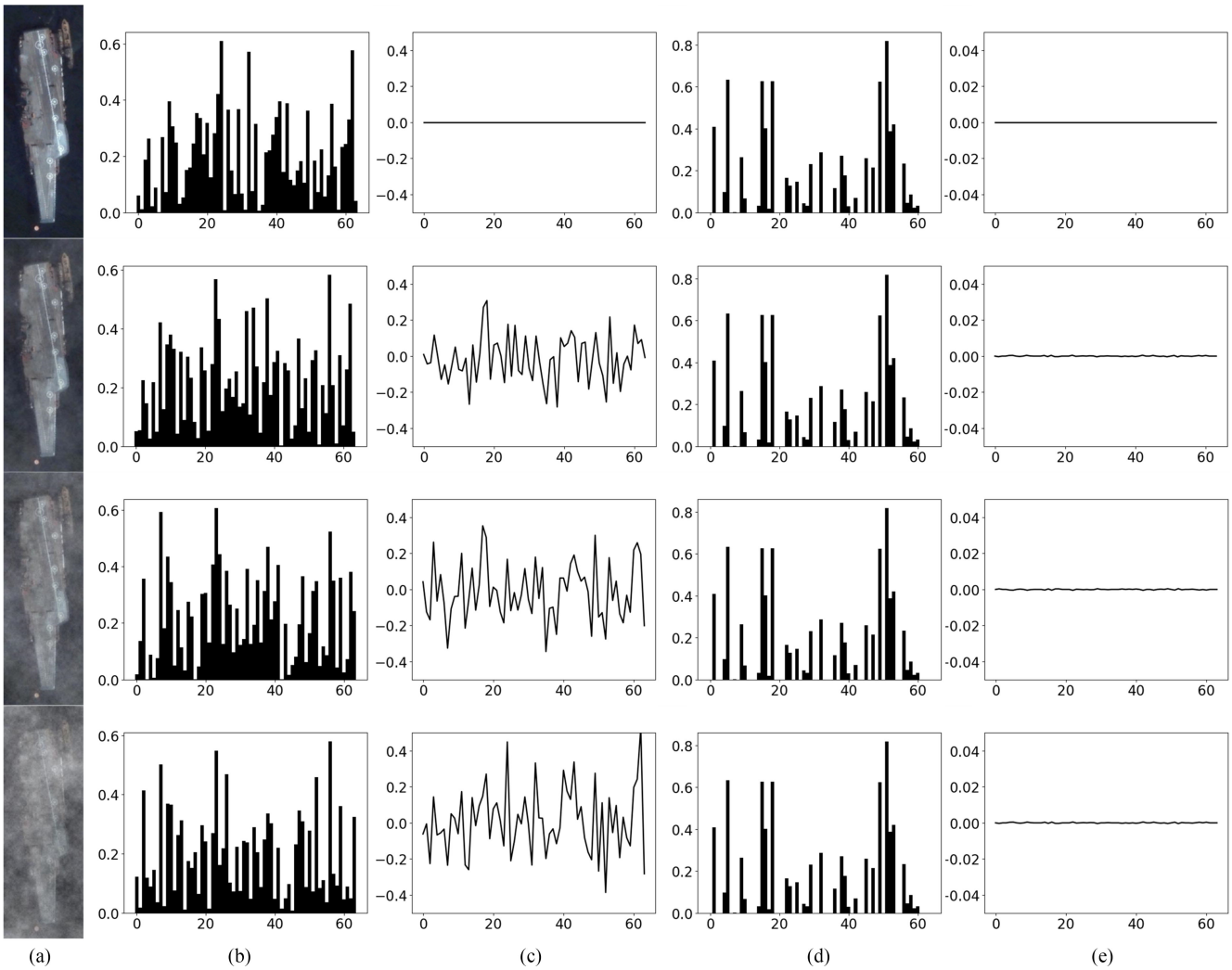

Fig. 4. Aircraft carrier characteristic histogram and feature line graph with occlusion levels of 0, 30, 50, and 70. (a) Input. (b) Inception V3 characteristic histogram. (c) Inception V3 characteristic difference line graph. (d) IICL-CNN characteristic histogram. (e) IICL-CNN characteristic difference line graph.

(PCA) [40]. The aircraft carrier in Fig. 4(a) is a sample with occlusion levels of 0, 30, 50, and 70, respectively. Fig. 4(b) shows the features of InceptionV3 and Fig. 4(d) shows the features of IICL-CNN. As the level of occlusion deepens, the features of the InceptionV3 network change markedly, while the features of the IICL-CNN remain stable. To analyze the transformation trend in more detail, the sample features with occlusion levels of 30, 50, and 70 are compared with the clear sample features with occlusion level of 0. The feature difference is illustrated by a line graph. As shown in Fig. 4(c) and Fig. 4(e), noticeable changes are represented in the different lines of InceptionV3, whereas IICL-CNN shows only minor oscillations. Overall, the features learned by IICL-CNN are highly similar between the clear samples and occlusion samples, which provide the basis for improving the accuracy and robustness.

B. Verification of the Constrained Objective Function

To further verify the effectiveness of the proposed feature similarity constraint in IICL-CNN, we randomly extracted 20 images with occlusion levels of 30, 40, 50, 60, and 70 from each of the nine categories. The average feature vector of the 100 samples of each category was then compared to that of the 100 clear images. Four metrics, including average value (Avg), standard deviation (Std), Interquartile Range (IQR), and signal-to-noise ratio (SNR) were derived from the average features with and without the feature similarity constraint, as shown in Table II. For all the nine categories, features with the constraint showed lower Avg, Std, and IQR, but higher SNR, indicating that the volatility of features was small when the similarity constraint was implemented.

C. Verification of the Center Loss

To verify whether center loss can reduce the distance between intraclasses, we randomly extracted 2000 samples from the test set of nine categories, and derived their features at the logits layer using InceptionV3 and IILC-CNN, respectively. To visualize the feature space, we reduced the dimension of the feature space to 2 using the t-distributed stochastic neighbor embedding (t-SNE) method [41]. Fig. 5(a) and (b) illustrates the reduced feature space of InceptionV3 and IICL-CNN, respectively. The feature

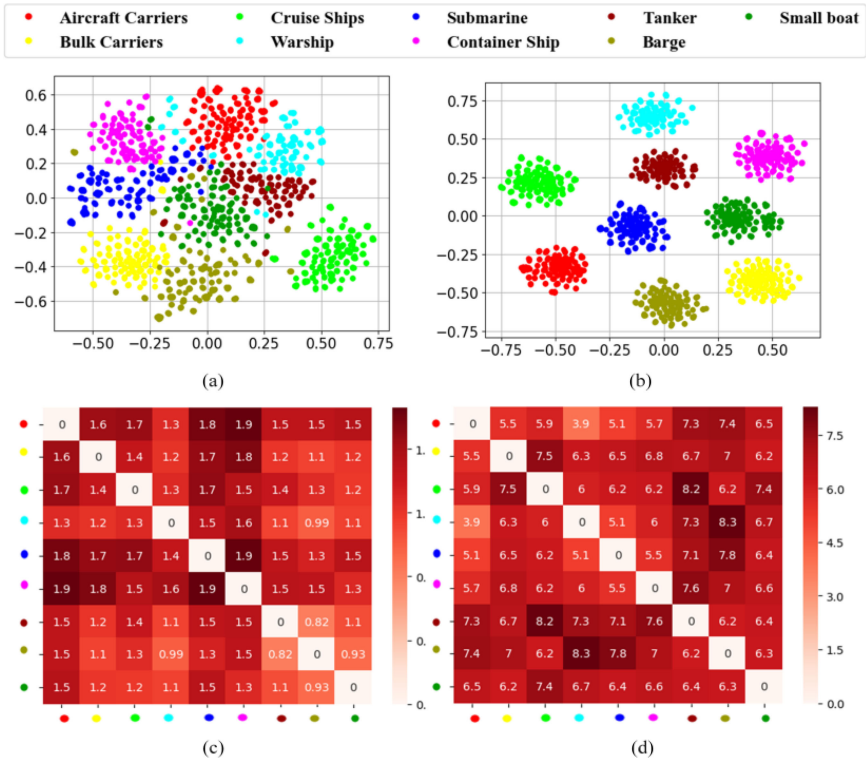

Fig. 5. Illustration of the clusters and the interclass distances using InceptionV3 and IICL-CNN features. (a) reduced feature space of InceptionV3. (b) reduced feature space of IICL-CNN. (c) inter-class distances of InceptionV3. (d) inter-class distances of IICL-CNN.

TABLE II
CONTRAST ERROR COMPARISON OF CATEGORY 9 SHIP FEATURES

type	constraint	Avg	Std	SNR	IQR
Aircraft carriers	Yes	0.196	0.632	4.552	0.844
	No	1.262	1.756	1.867	2.455
Bulk carriers	Yes	0.307	0.579	3.378	0.754
	No	1.149	1.540	1.196	2.012
Cruise ships	Yes	0.325	0.538	3.896	0.713
	No	1.591	1.770	1.767	2.295
Warship	Yes	0.208	0.485	5.700	0.637
	No	0.956	1.359	2.367	1.813
Submarine	Yes	0.097	0.427	7.047	0.572
	No	0.412	0.997	3.771	1.342
Container Ship	Yes	0.358	0.476	4.324	0.621
	No	1.312	1.235	1.549	2.911
Tanker	Yes	0.127	0.568	5.241	0.812
	No	1.236	1.538	3.154	2.135
Barge	Yes	0.214	0.543	4.675	0.651
	No	1.357	1.467	2.346	2.168
Small boat	Yes	0.176	0.611	3.894	0.771
	No	1.651	1.561	1.128	1.981

space shows that the clusters of the InceptionV3 features with no center loss are overlapped whereas, the clusters of the IICL-CNN features are more separable with higher compactness within each cluster.

We further calculated the mean Euclidean distances between the ship categories based on the InceptionV3 and IICL-CNN features to quantitatively evaluate the effectiveness of the center loss function. Fig. 5(c) and (d) shows the interclass distance between the ship categories using InceptionV3 features and IICL-CNN features, respectively. The interclass distances using IICL-CNN (6.53 ± 0.90) are much larger than those when using InceptionV3 (1.40 ± 0.26). Table III shows the intraclass distances within each ship category using the two models, and IICL-CNN demonstrates significantly lower intraclass distances than InceptionV3 ($p < 0.00001$). Higher interclass distances and lower intraclass distances imply that IICL-CNN has higher discriminating power.

D. Ship Target Classification

The target classification experiments were carried out using a workstation with an Intel i5-7300HQ (2.5 GHz) CPU and a NVidia GeForce GTX1060 GPU. During training, the clear samples and the corresponding occlusion samples were fed into the network for training as a batch with a small batch size of 6. The ratio between the clear samples and the occlusion samples is 1:5, and the occlusion samples include samples with the occlusion degrees of 30, 40, 50, 60, and 70. The total learning rate was set to 0.001 and the learning rate in center loss was set to 0.5. The models were trained with 30 000 iterations. We used grid search to identify the optimal combination of tradeoff coefficients of λ_1 and λ_2 for the cross-loss function and the feature similarity constraint. The tested values for both coefficients were $\{0.01, 0.005, 0.001, 0.0005\}$. Fig. 6 shows that when the parameter λ_1 was set to 0.001 and λ_2 was set to 0.01, the average accuracy was the highest.

Fig. 7 shows the accuracy and loss curves for the InceptionV3 and IICL-CNN. Compared with the InceptionV3 network,

TABLE III
INCEPTIONV3 AND IICL-CNN INTRACLASS DISTANCE WITHIN EACH SHIP CATEGORY

Type	Aircraft carriers	Bulk carriers	Cruise ships	Warship	Submarine	Container Ship	Tanker	Barge	Small boat
InceptionV3	1.632	1.339	1.485	1.313	1.712	1.476	0.845	1.042	1.188
IICL-CNN	**0.229**	**0.235**	**0.232**	**0.228**	**0.229**	**0.234**	**0.235**	**0.237**	**0.234**

Bold entities means the effect is better than other methods.

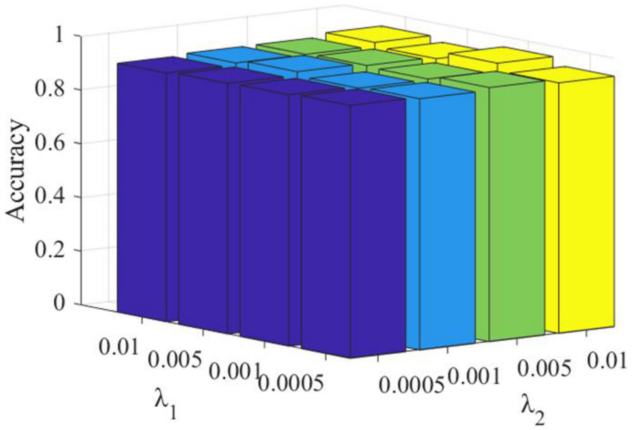

Fig. 6. Average accuracy of different parameters λ_1 and λ_2.

TABLE IV
AVERAGE ACCURACY OF EACH OBJECTIVE FUNCTION OF THE MODEL

Occlusion Level	L_l	L_c	L_l+L_c
0	95.62±0.11	95.42±0.14	**95.41±0.15**
30	91.27±0.17	91.05±0.23	**91.77±0.21**
40	88.73±0.21	87.25±0.18	**89.45±0.13**
50	83.61±0.21	79.73±0.23	**84.91±0.13**
60	81.18±0.12	74.68±0.11	**82.94±0.15**
70	78.95±0.14	69.73±0.12	**80.45±0.18**

Bold entities means the effect is better than other methods.

IICL-CNN had a faster convergence rate, higher accuracy, and more stable network as the number of iterations increased.

To test the role of L_l and L_c in the IICL-CNN model, L_l and L_c are trained separately and compared with joint training. The results are shown in Table IV. It can be seen that when the objective function is only L_l, the accuracy rate is not much different from the joint training, but when the objective function is only L_c, the accuracy rate drops rapidly after the occlusion degree is 50. Therefore, the contribution of the similarity constraint loss and center loss is validated.

To test the target classification performance of the IICL-CNN model, we compared it with InceptionV3 [14], InceptionV4 [15], SVDNet [25], PCANet [26], defogging algorithm [30], Resnet50 [42], and a traditional method using SVM classifier and the histogram of oriented gradients (SVM + HOG) [43]. The test set contains 335 clear ship images and their simulated fog occluded images at five occlusion levels resulting in a total number of 2010 samples. After training each network model,

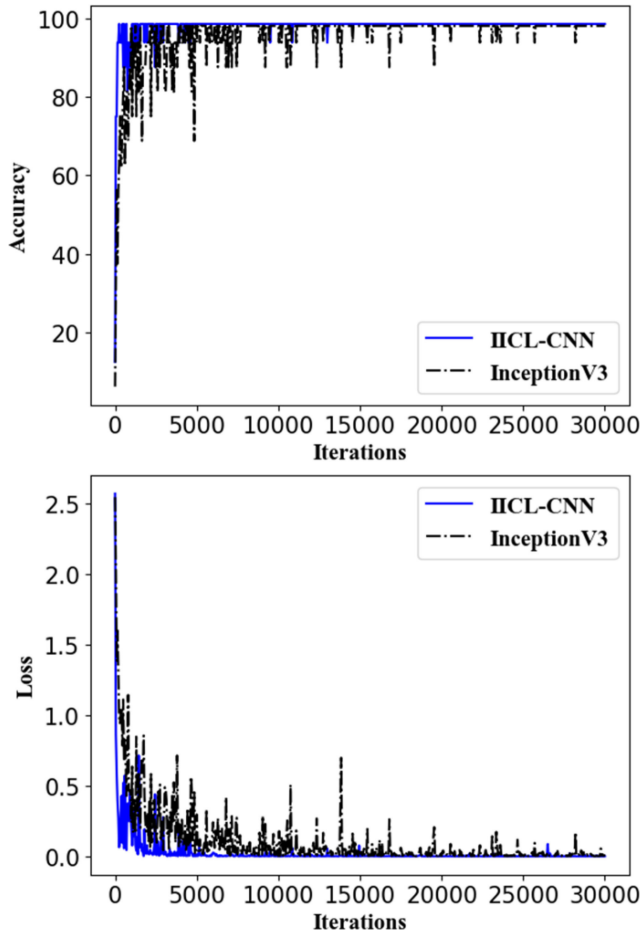

Fig. 7. Accuracy and loss curve for InceptionV3 and IICL-CNN networks.

test sets with different occlusion levels were used to calculate the average accuracy rate of each network model. As shown in Table V, the average accuracy of all models decreased as the level of occlusion increased. When the occlusion level was greater than 50, the average accuracy of InceptionV3, InceptionV4, and the traditional method were significantly reduced. IICL-CNN, SVDNet, and PCANet were robust to occlusion samples, and IICL-CNN achieved the best performance among all the compared algorithms.

The IICL-CNN model is used to test the accuracy of different ships under different degrees of occlusion. Table VI shows that as the degree of occlusion increases, the accuracy for all ship categories decreases. Also, similar shaped ship categories of

TABLE V
AVERAGE ACCURACY OF EACH NETWORK MODEL (%)

Occlusion Level	IICL-CNN	InceptionV3	InceptionV4	Resnet50	SVDNet	PCANet	Defogging	SVM+HOG
0	**95.41±0.15**	93.51±0.11	94.14±0.11	94.23±0.13	94.61±0.21	93.19±0.21	93.84±0.12	89.37±0.26
30	**91.77±0.21**	87.54±0.18	88.48±0.14	88.36±0.21	90.13±0.15	88.63±0.12	90.05±0.14	80.55±0.31
40	**89.45±0.13**	83.29±0.14	83.69±0.16	83.45±0.12	86.23±0.16	84.42±0.25	86.76±0.17	76.02±0.22
50	**84.91±0.13**	78.93±0.16	79.92±0.21	79.63±0.18	83.25±0.14	81.15±0.18	82.53±0.23	73.28±0.21
60	**82.94±0.15**	72.85±0.21	73.44±0.10	73.49±0.15	80.68±0.13	79.56±0.23	81.07±0.14	68.64±0.14
70	**80.45±0.18**	65.37±0.22	65.43±0.17	65.52±0.13	78.21±0.15	75.99±0.22	77.40±0.17	55.73±0.11

Bold entities means the effect is better than other methods.

TABLE VI
ACCURACY OF DIFFERENT SHIPS FOR EACH DEGREE OF OCCLUSION (%)

Class\ Occlusion	0	30	40	50	60	70
Aircraft carriers	96.77±0.12	93.38±0.21	91.21±0.15	86.32±0.11	83.25±0.17	81.33±0.11
Bulk carriers	94.32±0.23	90.31±0.11	89.51±0.10	84.36±0.16	81.36±0.13	78.37±0.11
Cruise ships	96.21±0.06	93.32±0.10	91.35±0.13	86.17±0.15	83.21±0.10	81.12±0.12
Warship	96.47±0.11	93.87±0.14	90.89±0.17	86.88±0.16	82.89±0.11	81.01±0.14
Submarine	97.11±0.14	94.21±0.12	91.47±0.11	86.13±0.17	83.28±0.16	80.68±0.14
Container Ship	94.56±0.12	90.47±0.16	88.32±0.12	83.98±0.17	81.02±0.12	78.87±0.13
Tanker	93.75±0.13	90.78±0.13	88.78±0.17	84.87±0.11	81.33±0.16	78.63±0.12
Barge	95.32±0.11	93.33±0.14	91.88±0.15	86.28±0.18	83.71±0.11	81.01±0.13
Small boat	96.11±0.08	94.11±0.09	91.41±0.10	86.38±0.11	83.25±0.08	81.37±0.13

TABLE VII
REAL-TIME CALCULATION TIME FOR EACH NETWORK MODEL (UNIT: S)

Model	InceptionV3	IICL-CNN	InceptionV4	SVDNet	PCANet	Defogging	SVM+HOG
Classification running time	8.322±0.012	8.432±0.031	9.012±0.018	9.126±0.041	8.303±0.024	8.573±0.017	6.465±0.021

bulk carriers, container ships, and tankers resulted in an accuracy lower than other ships.

In order to verify the time efficiency of the IICL-CNN model, we compared the calculation time of the model with InceptionV3, InceptionV4, SVDNet, PCANet, defogging algorithms, and traditional SVM + HOG methods. The calculated average classification running time is shown in Table VII. The SVM + HOG method has the shortest average classification time of 6.465 s, whereas IICL-CNN is about 30% slower (8.432 s).

Overall, IICL-CNN shows a good applicability in the field of maritime transportation with excellent accuracy, but longer calculation time than the SVM + HOG method, indicating there is room for further improvement in its real-time performance.

IV. DISCUSSION

Complex meteorological environment on the sea surface may lead to incomplete information of the ship targets and make them less recognizable. Fog occlusion is one of the common situations encountered in the sea environment. Ship targets that are obscured by the fog will pose changes to the ship target classification algorithms, thus leading to drop in accuracy. Previous effort to alleviate the impact of fog occlusion aims to remove the fog component from images, such as the defogging method proposed by Yoon *et al*. [30]. However, this method only uses the statistical color features and, therefore, is not able to cover the full spectrum of fog occlusion conditions. Furthermore, defogging, as a preprocessing step, is independent of the ship target classification model and thus cannot be trained in an end-to-end fashion like the CNN models.

We believe this challenge can be directly addressed by empowering the CNN model to learn the most robust and representative features of the occluded ship targets. We hypothesize that if the features are commonly applicable to both clear samples and obscured samples, then the accuracy for the occlusion samples will increase. To test our hypothesis, an IICL-CNN model was proposed based on the well-established inception network. The prominent improvement of IICL-CNN over InceptionV3 is that a FC layer was added to merge the features learned from samples with and without fog occlusion and to reduce the impact of feature divergence on target classification.

Furthermore, we added a similarity constraint and a center loss function to the objective function, which enforce the similarity between clear ship features and occlusion ship features, and

minimized the intraclass distances for further improvement of the features' discriminating power.

The InceptionV3 model was chosen in consideration of both performance gain and time cost. The performance of InceptionV3 and V4 are very similar according to our experiments as shown in Table V. The performance of InceptionV4 is slightly higher than that of InceptionV3, but when occlusion degree reaches 70, differences between the two become minimal (difference in accuracy < 0.0006). We also compared Resnet50 with InceptionV3, and similar pattern was noticed (difference < 0.0015). In terms of time cost, InceptionV4 almost doubles the training time compared to InceptionV3 (0.621 versus 0.316 s per step). InceptionV3 is also 30% faster than Resnet50 (0.411 s per step). To balance performance gain and time cost, we chose InceptionV3 as the basic network architecture in this study, and other networks were used as baseline models.

Compared to InceptionV3, the features learned by IICL-CNN are highly similar between the clear samples and occlusion samples as evaluated using the Avg, Std, IQR, and SNR metrics, which implies high robustness of the IICL-CNN features. Furthermore, the higher interclass distances and lower intraclass distances of the IICL-CNN features further imply higher discriminating power of the IICL-CNN model. In terms of the model's performance, as the level of occlusion increases, the difference in accuracy of the two models also increases, from 0.9% on clear samples to 17.48% on severely occluded samples (70% occlusion level). Compared to other state-of-the-art CNN-models and traditional feature extraction and classification method, our model demonstrates superior performance on classification of both the clear and occluded ship targets.

Another important contribution of this study is that we shaped a specific ship image dataset, which has a high potential to be used for benchmarking the state-of-the-art ship target classification algorithms. There are 48 000 ship images in this dataset across nine different ship categories covering various fog occlusion conditions. All images are standardized by removing the land component, resizing to 299×299 image size and applying zero-mean normalization. As the remote sensing images are becoming increasingly available, this dataset can be expanded to cover more samples and ship categories, which may further improve the ship target accuracy.

V. Conclusion

Ship target classification is a key technology in maritime traffic control, maritime target tracking, and military reconnaissance, etc. The complex ocean environment, including fog occlusion, often obscures the ship targets and poses challenges for ship target classification.

To improve the accuracy of the fog occluded ship targets, we propose the IICL-CNN model which refines the InceptionV3 structure and loss function to allow learning of similar features from both clear ship samples and obscured ship samples. Compared with the original InceptionV3, the average accuracy of the ship targets with 30%, 50%, and 70% fog occlusion levels are increased by 4.23%, 5.98%, and 17.48%, respectively. For the clear samples alone, the proposed IICL-CNN also demonstrates superior performance compared to other state-of-the-art CNN-models and traditional feature extraction and classification method. As the remote sensing images become increasingly available, we believe the performance of IICL-CNN can be further improved and will have a high applicability in tasks including maritime traffic control, onshore search, and maritime target tracking.

Acknowledgement

The authors would like to thank Mrs. Laya Jose for proofreading this Article.

References

[1] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010.

[2] Y. Xia, S. Wan, P. Jin, and L Yue, "A novel sea-land segmentation algorithm based on local binary patterns for ship detection," *Int. J. Signal Process. Image Process. Pattern Accuracy*, vol. 7, no. 3, pp. 237–246, 2014.

[3] J. Lu, Z. Xi, X. Yuan, G. Yu, and M. Zhang, "Ship target accuracy using high resolution range profiles based on FMT and SVM," in *Proc. IEEE CIE Int. Conf. Radar*, 2011, pp. 1299–1302.

[4] K. Ji, X. Leng, H. Wang, and S. Zhou, "Ship detection using weighted SVM and M-CHI decomposition in compact polarimetric SAR imagery," in *Proc. IEEE Int. Geosc. Remote Sens. Symp.*, 2017, pp. 890–893.

[5] P. Pan, Z. Jiang, J. Wu, H. Zhang, and P. Luo, "Ship accuracy based on active learning and composite kernel SVM," in *Proc. Chin. Conf. Image Graph. Technologies*, 2015, pp 198–207.

[6] Z. Li *et al.*, "Remote sensing ship target detection and accuracy system based on machine learning," *IGARSS 2019-2019 IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1272–1275.

[7] F. Bi, F. Liu, and L. Gao, "A hierarchical salient-region based algorithm for ship detection in remote sensing images," in *Advances in Neural Network Research and Applications*, Berlin, Germany: Springer, 2010, pp. 729–738.

[8] Y. Xia, S. Wan, and L. Yue, "A novel algorithm for ship detection based on dynamic fusion model of multi-feature and support vector machine," in *Proc. 6th Int. Conf. Image Graph.*, 2011, pp. 521–526.

[9] Z. Dan, N. Sang, R. Wang, Y. Chen, and X. Chen, "A transductive transfer learning method for ship target accuracy," in *Proc. 7th Int. Conf. Image Graph.*, 2013, pp. 418–422.

[10] S. Tong, K. Sun, B. Shi, and J. Chen, "A ship target automatic accuracy method for sub-meter remote sensing images," in *Proc. 4th Int. Workshop Earth Observ. Remote Sens. Appl.*, 2016, pp. 153–156.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[12] G. Cheng, P. Zhou, and J. Han. "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[13] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Accuracy*, 2015, pp. 1–9.

[14] C. Szegedy, V. Vanhoucke, S. Loffe, J. Shlens, and Z. Wojna. "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern*, pp. 2818–2826, 2016.

[15] C. Szegedy, S. Loffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[16] A. H. Tuttle *et al.*, "A deep neural network to assess spontaneous pain from mouse facial expressions," *Mol. Pain*, vol. 14, 2018, Art. no. 1744806918763658. [Online]. Available: http://dx.doi.org/10.1177/1744806918763658

[17] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, "Comparison of transferred deep neural networks in ultrasonic breast masses discrimination," *BioMed Res. Int.* vol. 2018, 2018, Art. no. 4605191.

[18] S. Jain and J. Dhar. "Image based search engine using deep learning," in *Proc. 10th Int. Conf. Contemporary Comput.*, 2017, pp. 1–7.

[19] R. D. Samuel and B. R. Kanna. "Tuberculosis (TB) detection system using deep neural networks," *Neural Comput. Appl.*, vol. 31, no. 5, pp. 1533–1545, 2019.
[20] X. Hou, Q. Xu, and Y. Ji, "Ship detection from optical remote sensing image based on size-adapted CNN," in *Proc. 5th Int. Workshop Earth Observ. Remote Sens. Appl.*, 2018, pp. 1–5.
[21] Z. Liu, J. Hu, L. Weng, and Y. Yang, "Rotated region based CNN for ship detection," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 900–904.
[22] J. Sun, H. Zou, Z. Deng, X. Cao, M. Li, and Q. Ma, "Multiclass oriented ship localization and accuracy in high resolution remote sensing images," in *Proc. IGARSS 2019-2019 IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1288–1291.
[23] F. Lei, W. Wang, and W. Zhang, "Ship extraction using post CNN from high resolution optical remotely sensed images," in *Proc. IEEE 3rd Inf. Technol. Netw. Electron. Automat. Control Conf.*, 2019, pp. 2531–2535.
[24] Q. Zhang, J. Lu, T. Liu, P. Zhang, and Q. Liu, "Ship HRRP target accuracy based on CNN and ELM," in *Proc. 4th Int. Conf. Electromechanical Control Technol. Transp.*, 2019, pp. 124–128.
[25] Z. Zou, and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Trans. Geosc. Remote Sens.*, vol. 54, no. 10, pp. 5832–5845, Oct. 2016.
[26] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.
[27] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.
[28] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.
[29] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 860.
[30] I. Yoon, J. Jeon, J. Lee, and J. Paik, "Weighted image defogging method using statistical RGB channel feature extraction," in *Int. SoC Des. Conf.*, 2010, pp. 34–35.
[31] B. Yao, L. Huang, and C. Liu, "Adaptive defogging of a single image," in *Proc 2nd Int. Symp. Comput. Intell. Des.*, 2009, pp. 56–59.
[32] H. Park, J. Park, H. Kim, and J. Paik, "Improved DCP-based image defogging using stereo images," in *Proc. IEEE 6th Int. Conf. Consum. Electron.*, 2016, pp. 48–49.
[33] C. Li, T. Fan, X. Ma, Z. Zhang, H. Wu, and L. Chen, "An improved image defogging method based on dark channel prior," in *Proc. 2nd Int. Conf. Image, Vis. Comput.*, 2017, pp. 414–417.
[34] I. Yoon, M. H. Hayes, and J. Paik, "Wavelength-adaptive image formation model and geometric classification for defogging unmanned aerial vehicle images," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2013, pp. 2454–2458.
[35] W. Dong, X. Zhang, and C. Zhang, "Generation of cloud image based on Perlin noise," in *Proc. Int. Conf. Multimedia Commun.*, 2010, pp. 61–63.
[36] C. Sun, B. Kong, L. He, and Q. Tian, "An algorithm of imaging simulation of fog with different visibility," *IEEE Int. Conf. Inf. Automat.*, 2015, pp. 1607–1611.
[37] W. Huang and Y. Tang. *Tensorflow Combat*. Beijing, China: Publishing House of Electronics Industry, 2017, p. 142.
[38] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization," *Comput. Sci.*, vol. 4, no. 3, pp. 616–624, 2014.
[39] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face accuracy," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 499–515.
[40] H. Abdi and L. J. Williams. "Principal component analysis," *Wiley Interdisciplinary Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[41] L. Maaten and G. Hinton. "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[43] B. Alsahwa, F. Maussang, R. Garello, and A. Chevallier, "Marine life airborne observation using HOG and SVM classifier," in *Proc. OCEANS MTS/IEEE Monterey.*, 2016, pp. 1–5.

Kun Liu received the B.S. degree in automation from Jilin University, Changchun, China, in 2004, and the M.S. and Ph.D. degrees in pattern recognition and intelligent system from Northwestern Polytechnical University, Xi'an, China, in 2007 and 2009, respectively.

She is currently an Associate Professor with the School of Information Engineering, Shanghai Maritime University, Shanghai, China. Her main research interests include computer vision, image processing, and artificial intelligence.

Shengtao Yu received the B.Sc. degree in communication engineering from the Qiqihar University, Qiqihar, China. He is currently working toward the M.S. degree in communication and information system with the Shanghai Maritime University, Shanghai, China.

His main research interests include image processing and deep learning.

Sidong Liu was born in Heilongjiang Province, China, in 1984. He received the B.Eng. degree in biological information technology from the Harbin Institute of Technology, Harbin, China, in 2007, and the M.App.Sc. degree in bioinformatics, the M.IT degree in computer science, and the PhD degree in computer science from the University of Sydney, Sydney, NSW, Australia, in 2009, 2010 and 2015, respectively.

He is currently an NHMRC Early Career Fellow with the Australian Institute of Health Innovation, Macquarie University, Sydney, NSW, Australia. His research interests include deep learning, image processing, and medical informatics.