Maritime Visible Image Classification Based on Double Transfer Method

RUIYU ZHAO, (Student Member, IEEE), JIANHUA WANG, XIANG ZHENG, JING WEN, LIUZHONG RAO, AND JUNXIA ZHAO

Key Laboratory of Marine Technology and Control Engineering, Shanghai Maritime University, Shanghai 201306, China

Corresponding author: Ruiyu Zhao (zhaoruiyu@stu.shmtu.edu.cn)

ABSTRACT

Image classification using deep transfer learning has received significant attention, benefiting from pre-trained with the large-scale annotation dataset and continuous improvement of neural network structure. In contrast to universal image classification, however, few publicly available datasets of maritime environments utilize deep transfer learning. Due to data-gathering effort and computational cost, the maritime datasets are deficient in the method of merging datasets and the benchmark of few-shot dataset classifiers. In this article, we proposed the double transfer method, consisting of the merging datasets network and the backbone network, to address the problem. The merging datasets network measuring image similarity separates classes of known and unknown samples to reorganize a dataset, and the backbone network is constructed from the model EfficientNet-b5 by network-based deep transfer learning. Using the merging datasets network, we introduce the visible maritime image dataset, which has 3,750 images and twenty-five classes, including multitudinous maritime objects. The backbone networks evaluated and analyzed the dataset based on accuracy, precision, recall, and F-measure metrics. Using the double transfer method, we can achieve an accuracy of 91.39% in the visible maritime image dataset.

INDEX TERMS

Image classification, maritime environments, double transfer method, merging datasets, deep transfer learning.

I. INTRODUCTION

In the last decades, the application of computer vision in maritime environments can monitor marine traffic and sea border activities, detect the ordinary ship with illegally occupying individual channels and anchorage grounds, and collect intelligence on illegal fishing and smuggling [1]–[3]. Large-scale annotated datasets train convolutional neural networks (CNNs) with supervised learning to achieve image classification in recent years [4]. However, the process of data training suffers considerable data-gathering effort and computational cost, and publicly available datasets are scarce in maritime environments [5], [6]. There are some maritime datasets, such as object detection and tracking [7], piracy detection [8], and obstacle detection [9], [10], but none of them is used for maritime classification [6]. Compared to the remote sensing datasets and infrared (IR) datasets [2], the electro-optical visible dataset is insufficiently applied to maritime environments [6], [7]. The majority of maritime visible classification datasets is closed source to difficultly obtain, such as the Maritime Detection Classification and Tracking (MarDCT) datasets [11] and E2S2-Vessel dataset [12]. Besides, rare categories are hard to collect, and a few instances toughly train a deep CNNs of a satisfying accuracy. The visible and infrared ship imagery (VAIS) coarse-grained dataset [13] and the Maritime VEsseLs (MARVEL) fine-grained dataset [14] are the few publicly available datasets in the maritime image classification. However, the class imbalance appears in the datasets, which affect the result of image classification [6], [15]. In contrast to the two publicly available datasets, few datasets merge different existing annotated maritime datasets and unlabelled data in maritime environments for image classification.

The state-of-the-art CNNs using deep learning achieved significant success in image recognition, which are the effective models to tackle a variety of visual tasks [4], [16], [17]. However, the CNN application of the ship classification in maritime environments began in 2015 [13]. Due to the shortage of well-labeled dataset and robust algorithm or specific benchmark process, the majority of solutions dwell on coarse-grained classification instead of
fine-grained or instance-level image classification [11], [14], [18], [19]. Numerous approaches remain the design and improved CNN model based on AlexNet and VGGNet [4], [12]–[14], [18]–[21]. The state-of-the-art CNN model, such as EfficientNet, is rarely used in the ship classification with the VAIS dataset or the MARVEL dataset [22].

The main contributions of this article are summarized as follows: (1) The double transfer method was proposed, consisting of the merging datasets network and the backbone network. The merging datasets network measuring image similarity reorganize disparate datasets to obtain fine-grained datasets. Then, it is used for the Visible Maritime Image (VMI) dataset, which has 3,750 images and twenty-five class. The network eliminates class imbalance [15], and improve similarity in the same class. (2) Compared with the elementary convolutional neural network, the backbone network is structured with the model EfficientNet-b5 using deep transfer learning, and draw a comparison with AlexNet, VGGNet, ResNet, DenseNet, SqueezeNet, and other EfficientNet backbone networks. The backbone networks evaluated and analyzed the VMI dataset based on accuracy, precision, recall, and F-measure metrics.

The rest of this article is organized as follows. Section II reviews the maritime dataset and image classification. Section III introduces the double transfer method consisting of the merging datasets network and the backbone network. Section IV describes the VMI datasets by the merging datasets network. Section V reports experiments and results with the VMI datasets using the backbone network. Finally, the conclusion and discussion are given in Section VI.

II. RELATED MARITIME DATASETS

In the maritime image classification, the commonly available datasets are the VAIS dataset and MARVEL dataset.

A. THE VAIS DATASET

The VAIS dataset consists of 1,623 visible images in six coarse-grained categories [13], consisting of merchant ship, sailing ship, medium passenger ship, medium “other” ship, tugboats, and small boat. To analyze the quality of the dataset it is necessary to check if the classes are equally distributed [6]. Fig. 1 shows the distribution of “official” training and test set, which partition the VAIS dataset into about (54%, 46%) ratio of six coarse-grained categories [18], [19]. It is indicated that The VAIS data have a strong class imbalance. Another important property for image classification is image resolution. The majority of image resolution is less than 50,176 (224 × 224) pixels which is less than the input resolution of lots of CNNs, and 44 images have insufficient 1,024 (32 × 32) pixels with losing local details which ultimately results in lower performance of image classification [2], [4], [17], [20], [22]–[25].

Zhang et al. [13] presented the gnostic field and CNN that has been pre-trained on ImageNet, and use it to extract features from VAIS for image classification to achieve the best accuracy of 82.4% with only using the gnostic field.

However, the fine-tuning of a pre-trained model yields poor performance, an accuracy of 81.9%, which less than 82.4%. Khellal et al. [26] proposed the problem of overfitting due to the lack of training data, even when using regularization techniques. Shi et al. [19] proposed the method consists of a multi-feature ensemble based on AlexNet CNN (ME-CNN), and use a two-dimensional discrete fractional Fourier transform (2D-DFrFT) with amplitude (M) and phase (P) to extract information. Using the ME-CNN with training from scratch achieves the best accuracy of 87.33%. Then, Fine-tuning CNN with multi-scale rotation invariance features is proposed to achieve an accuracy of 88.00% [18].

B. THE MARVEL DATASET

The MARVEL dataset is a large-scale, fine-grained image dataset, including 26 superclasses and 239,622 images. The dataset has an equal number of examples in each superclass from official that the proportion of the training set and the test set is 8:1 (88.89% and 11.11%) [14]. All images of the dataset is greater than 155,236 (approximately 394 × 394) pixels. The critical property of fine-grained image classification is vessel taxonomy. Statcode5 ship coding system proposed by IHS Markit is a standard ship type that is easier to see a logical breakdown and allocation [3], [27]. The image of MARVEL is collected from website shipspotting [28]. However, the superclasses of the dataset contain the ship and vessel, but no other maritime targets are considered, such as boat and buoy, shown in Table 1. Besides, the statistics show that of all those classes, the followings are the dominant ones (over 80%) in terms of the number of images in reality: cargo ship, tanker, bulk carrier, container ship and passenger ship [12], [14].

Gundogdu et al. [14] proposed the application of AlexNet to achieve an accuracy of 73.15%. However, using SVM and pre-trained VGGNet only obtain an accuracy of 53.89%, and few types, such as livestock carrier, fire fighting vessel, and training ship, have only a precision of 20%-50%. Leclerc et al. [3] proposed the application of GoogLeNet to achieve an accuracy of 78.73% in the MARVEL dataset.

Besides, some data prove that training from scratch can achieve better experimental results [2]. He et al. [29] believes that proper pre-training can make up for the lack of data to bring a significant improvement in the case of small data sets and insufficient computing power and shorten the research
of task said to learn from experience speaking, machine learning is a computer program, which contains a variety of state-of-the-art CNNs models. Generally, Deep Learning is a branch of machine learning, which contains only a state-of-the-art CNN model. A task $T$ is a particular solution of deep learning [5]. However, deep learning contains only a limited number of examples with supervised information for the target domain(s). The number of fine-tuning layers has no apparent relationship with the accuracy. Donahue et al. [30] considers that model A is trained with the training set A-Data, and fine-tuning can reach a level similar to the original on the dataset B-Data. The number of fine-tuning layers has no apparent relationship with the accuracy. Donahue et al. [31] showed that features extracted from a deep CNN trained on a large-scale dataset could be reused for new generic tasks.

III. DOUBLE TRANSFER METHOD

Deep Learning is a branch of machine learning, which contains a variety of state-of-the-art CNNs models. Generally speaking, machine learning is a computer program, which is said to learn from experience $E$ with respect to some classes of task $T$ and performance measure $P$ if its performance can improve with $E$ on $T$ measured by $P$. However, deep learning is trying to obtain perfect performance due to restricted to the few-shot data and insufficient model training. Few-shot learning (FSL), which is a type of machine learning problem (specified by $E$, $T$, and $P$), where $E$ contains only a limited number of examples with supervised information for the target $T$, is a particular solution of deep learning [5]. Few-shot and fine-grained dataset is the FSL problems, which is difficult from the classification of a large-scale dataset. The double transfer method consists of two aspects that one is merging multiple data sources, and another is the adjustment of the CNNs model with network-based transfer learning to address the problem. First of all, the notation and definition are shown in the following.

A. NOTATION AND DEFINITION

The double transfer method is based on transfer learning, whose most important concepts are domain and task. A domain $D$ is composed of two parts, i.e., a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, i.e., $D = \{\mathcal{X}, P(X)\}$. $P(X)$ is hard to obtain its explicit formulation. The symbol $X$ denotes an instance set, which is defined as $X = \{x_i | x_i \in \mathcal{X}, i = 1, \cdots, n\}$. A task $T$ consists of a label space $\mathcal{Y}$ and the target prediction mapping $f : \mathcal{X} \rightarrow \mathcal{Y}$, i.e., $T = \{y, f(x)\}$ [32], [33].

Therefore, the definition of transfer learning can be obtained. Given some/an observation(s) corresponding to $m_S \in \mathbb{N}^+$ source domain(s) and task(s) (i.e., $\{(D_{S_i}, T_{S_i}) | i = 1, \cdots, m_S\}$), and some/an observation(s) about $m_T \in \mathbb{N}^+$ target domain(s) and task(s) (i.e., $\{(D_T, T_T) | i = 1, \cdots, m_T\}$), transfer learning utilizes the prior knowledge of experience $E$ implied in the source domain(s) to improve the performance of the learned decision functions $f_T(j = 1, \cdots, m_T)$ on the target domain(s) [5], [32]. Then, there is the definition of deep transfer learning. Given a transfer learning task defined by $\{(D_S, T_S, D_T, T_T, f_T)\}$. It is a deep transfer learning task where $f_T(\cdot)$ is a non-linear function that reflected a deep neural network [33].

However, we can only take a finite number of samples from the domain $D$. So, all hypothesis mappings $h : X \rightarrow Y$ make up hypothesis space $\mathcal{H}$. The effect of deep transfer learning on the hypothesis space is shown in Fig. 2. In the following discussion, only a limited sample is considered.

![FIGURE 2. The prior knowledge of deep transfer learning affects the hypothesis mappings $h$. The whole area is knowledge, including unknown knowledge and known knowledge, which is the mapping of the domain $D$ on the hypothesis space $\mathcal{H}$. Non-white areas are hypothetical space $\mathcal{H}$. The blue area is prior knowledge space, which can reduce the search range of $h$. CNN model is the gray area $\mathcal{H}_{opt}$. Point start is a search starting point. Optimizer, such as Adam optimizer [34], is the search path. $h_b$ is empirical best, which is the achievable hypothesis mapping. $h^*$ is best in $\mathcal{H}_{opt}$, which demands the state-of-the-art CNN model. $h$ is optimal mappings, however, which demands the nonexistence of unknown knowledge. Deep transfer learning using prior knowledge reduce restricted factors for the best $h^*$.](image-url)

In the classification, given the sample set is defined as $D_C = \{x_1, \cdots, x_i\}, i \in [1, N]$, $N$ is the total number of samples. Assuming the label $y_i \in [1, \cdots, c], c \leq N$ and the mapping hypothesis $h(D_C)$ that is trained, the predicted set is $T_C = \{(x_1^{(c)}, y_1), \cdots, (x_i^{(c)}, y_i)\}$ the correct predicted label. $X_C = \{x_i^{(c)} | x_i^{(c)} \in \mathcal{X}, i \in [1, N]\}$ is the extracted feature via $h(D_C)$, and $x_i^{(c)}$ defined as

$$x_i^{(c)} = [g, u_i, s_i]^T$$

TABLE 1. Comparison of collection methods used in maritime visible image classification [5].

| Class No. | MARVEL subclasses | Staicoder’s ship coding system |
|-----------|------------------|-------------------------------|
| 1         | Container Ship   | Container Ship                |
| 2         | Bulk Carrier     | Bulk Carrier                  |
| 3         | Passengers Ship  | Passengers Ship               |
| 4         | Ro/ro-passerger Ship | Passenger/Ro-Ro Cargo Ship |
| 5         | Ro-Ro Cargo      | Ro-Ro Cargo Ship              |
| 6         | Bulk           | Bulk                          |
| 7         | Vehicles Carrier | Vehicles Carrier              |
| 8         | Reefer          | Refrigerated Cargo Ship       |
| 9         | Yacht           | Leisure Vessels               |
| 10        | Sailing Vessel  | Sailing Vessel                |
| 11        | Heavy Load Carrier | Heavy Load Carrier            |
| 12        | Wood Chips Carrier | Wood Chips Carrier            |
| 13        | Livestock Carrier | Livestock Carrier            |
| 14        | Fire Fighting Vessel | Fire Fighting Vessel        |
| 15        | Patrol Vessel   | Patrol Vessel                 |
| 16        | Platform        | Platform                      |
| 17        | Standby Safety Vessel | Standby Safety Vessel       |
| 18        | Combat Vessel   | Combat Vessel                 |
| 19        | Training Ship   | Training Ship                 |
| 20        | Icebreaker      | Icebreaker                    |
| 21        | Replenishment Vessel | Replenishment Vessel     |
| 22        | Tankers         | Tankers                       |
| 23        | Fishing Vessels | Fishing Vessels               |
| 24        | Supply Vessels  | Offshore Tug/Supply ship      |
| 25        | Carrier/Floating | General Cargo Ship            |
| 26        | Dredger         | Dredger                       |

VOLUME 8, 2020 1665357
and $g, u_i, s_i$ are defined as
\[
g = (g_1, \ldots, g_m, 0_{m+1}, \ldots, 0_k)
\]
\[
u_i = (u_{i,1}, \ldots, u_{i,Cn}, 0_{Cn+1}, \ldots, 0_k)
\]
\[
s_i = (s_{i,1}, \ldots, s_{i,n}, 0_{n+1}, \ldots, 0_k)
\]
where, $1 < m, c_n, n \leq k$; $g$ is the identical feature in all classes $C$; $u_i$ is the similar feature in the correlation class $c_n$, $s_i$ is the special feature in each class. $g$ and $u_i$ constitute prior knowledge. Actually, due to $g$ describe the feature of all classes, (1) can be simplified to
\[
x_i^C = [u_i, s_i]^T
\]
(Meanwhile, the aim of the classifier find optimal results which is no correlation between the $x_i^C$. In other word, $u_i$ approaches zero, $\lim_{N \to \infty} u_i \rightarrow 0$, $i \in [1, N]$. However, the number of samples is finite. Given $U = [u_1, \ldots, u_i]$, there is $0 < \epsilon, \delta < 1$, defined as
\[
P(E(||U||_2) \leq \epsilon) \geq 1 - \delta
\]
where $||U||_2$ the L2 norm of $U$. With the classifier, there is at least a probability of $(1 - \delta)$ to learn the correct classes (The error is at most $\epsilon$) [35]. However, the feature $U$ is hardly obtained directly. We use the accuracy $ACC$ from the confusion matrix instead of $||U||_2$ to measure the association between classes. The confusion matrix is defined as

|     | Prediction |
|-----|------------|
|     | Positive   | Negative  |
| True| TP         | FN        |
|     | FP         | TN        |

where $TP = \sum_{i=1}^{N} I(y_i = h(x_i) = c)$, $FN = \sum_{i=1}^{N} I(y_i = c \land h(x_i) \neq c)$, $FP = \sum_{i=1}^{N} I(y_i \neq c \land h(x_i) = c)$, and $TN = \sum_{i=1}^{N} I(y_i \neq c \land h(x_i) \neq c)$. The accuracy $ACC$ is defined as
\[
ACC = (TP + TN)/(TP + FN + FP + TN).
\]
Therefore, (4) can be redefined as
\[
P(E(1 - ACC) \leq \epsilon) \geq 1 - \delta,
\]
where $P(\cdot)$ denotes the unknown probability distribution. Commonly, the only correct $ACC$ is used to measure the classification ability of the network.

Then, precision represents the proportion of correct positive samples predicted for all positive samples in the same class $c$, defined as
\[
P = (TP)/(TP + FP).
\]
Recall represents the proportion of correct predictions in all samples whose true label is class $c$, defined as
\[
R = (TP)/(TP + FN).
\]
F measure is harmonic mean with precision and recall, defined as
\[
F = ((1 + \beta^2) \cdot P \cdot R)/(\beta^2 \cdot P + R),
\]
where $\beta$ measures the relative importance of $R$ to $P$. Generally, $\beta = 1$ is selected, and $F$ measure is called $F$-1. Precision, recall, and $F$ measure a single class of metrics.

B. MERGING DATASET NETWORK
Source domain consists of the different dataset with dissimilar labeled system, and target domain have non-labeled samples. Given a set of domain $D_{S1} = \{x_{di}, y_{di}\}_{d_i=1}^{N_{S1}}$, $y_{di} \in \{1, C_i\}$, $D_{S2} = \{x_{dj}, y_{dj}\}_{d_j=1}^{N_{S1}}$, $y_{dj} \notin \{1, C_i\}$, and $D_{T1} = \{x_{dk}\}_{d_k=1}^{N_{T1}}$, there are two new domain, source domain $D_{S} = D_{S1} \cup D_{S2}$ and target domain $D_{T} = D_{T1}$, to acquire the source domain with labeled samples and the target domain without unlabeled samples. At least each class $c$ have ten sample $\sum_{c}x_{ic} \geq 10, c \in C_i$ in the source domain.

The merging datasets include a feature extractor $G_i$ and a classifier $G_{hi}, i \in \{1, 2\}$. $G_i$ is trained with source sample $(x_i, y_i)$ based on a pre-trained AlexNet model using deep transfer learning. $G_{h1}$ is $C_i$ classifier, and $G_{h2}$ is $(C_i + O_i)$ classifier, where the $O_i$ class has only one label “outlier”. Thus, the loss $L_{train}$ of the trained feature extractor and classifier $G_{hi}(G_i(x_i))$ is defined as:
\[
L_{train} = \frac{1}{N_{S}} \sum_{x_i} (G_{hi}(G_i(x_i)) \cdot \ln y_i
\]
\[
+ (1 - G_{hi}(G_i(x_i))) \cdot \ln (1 - y_i)).
\]
The outputs $w_i^S = \{p_i^1, \ldots, p_i^j\}$ of $G_{hi}(G_i(x_i))$ are the probability of that sample belonging to known class $c$. So, we use the average probability $w_i^S$ to measure similarity.
between target sample $x_j$ and source class, defined as

$$\min_c ||\bar{w}_c^S - w_j^T||^2_2,$$  \hspace{1cm} (11)

where $\bar{w}_c^S$ is the average probability matrix of class $c$ in the source domain. $w_j^T$ is outputs of $G_{h2}(G_t(x_j))$. The loss $L_{test}$ of the tested $G_{h2}(G_c)$ is defined as:

$$L_{test} = -\frac{1}{N_T} \sum_{x_j} [G_{h2}(G_t(x_j)) \ln y_j + (1 - G_{h2}(G_t(x_j))) \ln (1 - y_j)].$$  \hspace{1cm} (12)

Inputting $x_j$ to $G_{h2}(G_c)$, we define the similarity with two target sample which is

$$\min_c ||\bar{w}_c^S - w_j^T||^2_2 \leq \min_c ||\bar{w}_c^S - w_{j-1}^T||^2_2$$  \hspace{1cm} (13)

where (13) measure the maximum similarity that target samples $j$ is greater than target samples $j-1$. The current sample $x_j$ is labeled to $(x_j, y_j)$, and $\bar{w}_c^S$ changes with an increase in the labeled data, shown in Algorithm 1 row 10.

After the complete classification of the target sample, a few classes contain more samples than the sample maximum $N_c$. Therefore, the minimum similarity value in all samples are compared to find it as an outlier class. Finally, we obtain the new dataset $D_N = \{x_n, y_n\}_{n=1}^N$. The algorithm is shown in Algorithm 1. Besides, $N_t$ is the number of the target labeled sample in Algorithm 1. $N_t$ is initialized to $N_t = 0$, and $N_t$ only record classes $C_t$ rather than classes $(C_s + O_s)$.

C. THE BACKBONE NETWORK

We proposed the backbone network using network-based deep transfer learning. Network-based deep transfer learning refers to the reuse of the partial network that pre-trained in the source domain, including its network structure and connection parameters, transfer it to be a part of deep neural network which used in the target domain [33].

First, a large-scale dataset ImageNet is used to train the backbone network $G_{h2}(G_{p1}(G_t))$ to obtain optimal parameters. $G_{h2}$ is a source domain classifier. $(G_{p1}(G_t))$ is a feature extractor. $G_t$ is part of the network to be transferred. Then, the first few layers $G_t$ of the network are transferred to be a part of the transferred backbone network $G_{h2}(G_{p2}(G_t))$, where $G_{h2}$ is a target domain classifier, and $(G_{p2}(G_t))$ is a feature extractor. Finally, the target domain trains $G_{h2}(G_{p2}(G_t))$, shown in Fig. 3. The loss of the trained feature extractor $(G_{p2}(G_t))$ is defined as

$$L_t = \frac{1}{N_n} \sum_{i=1}^{N_n} \mathcal{L}(G_{h2}(G_{p2}(G_t(x_n))), y_n)$$  \hspace{1cm} (14)

where $\mathcal{L}$ is cross-entropy loss.

In the method, we have chosen the backbone network with the EfficientNet model, which is the state-of-the-art network structure in image classification. The backbone network using the EfficientNet model compares with AlexNet, VGGNet, ResNet, DenseNet, and SqueezeNet backbone networks. The structure of the proposed CNNs is shown in Fig. 4.

AlexNet is first used a large-scale dataset to train CNN [4], seen in Fig. 4(a). Compared to other networks, VGGNet has the largest network parameters and occupies more memory with higher complexity to improve accuracy [20], [25]. The VGGNet-11, VGGNet-16 are chosen in this article, seen in Fig. 4(b). SqueezeNet uses model compression to achieve AlexNet-level accuracy on ImageNet, reducing parameters by 50 times, and proposes a Fire structure [17], shown in Fig. 4(e)(i). ResNet introduces deep residual learning framework, which is

$$F(x) := H(x) - x$$  \hspace{1cm} (15)

where $H(x)$ is the desired underlying mapping, with $x$ denoting the inputs to the first of these layers to these layers approximate a residual function $F(x)$ [23], seen in Fig. 4(c)(g). DenseNet is built based on ResNet [24], shown in in Fig. 4(d)(h). Therefore, ResNet-18, ResNet-50, ResNet-101, DenseNet-121 and DenseNet-169 are chosen in this article.

Besides that, EfficientNet can provide higher accuracy by adjusting the network depth $d$, width $w$, and resolution $r$. EfficientNet-b5 and EfficientNet-b7 use a compound scaling method to scale up from EfficientNet-b0, shown in Fig. 4(f)(j). The compound scaling method is defined as

$$\begin{align*}
\text{depth} & : d = \alpha^\phi \\
\text{width} & : w = \beta^\phi \\
\text{resolution} & : r = \gamma^\phi \\
\text{s.t.} & : \alpha \cdot \beta^2 \approx 2 \\
& : \alpha \geq 0, \beta \geq 1, \gamma \geq 1
\end{align*}$$  \hspace{1cm} (16)

where $\alpha, \beta, \gamma$ are constants that can be determined by a small grid search. $\phi$ is a user-specified coefficient that controls how many more resources are available for model scaling, while $\alpha$, $\beta$, $\gamma$ specify how to assign these extra resources to network depth $d$, width $w$, and resolution $r$ respectively [22].

IV. VISIBLE MARITIME IMAGE DATASET

The merging dataset network is used to create the Visible Maritime Image dataset. The class of the VMI dataset
refers to the classified categories of the SMD, the VAIS dataset, the MARVEL dataset, the naval ship dataset, and Statcode 5 Shiptype coding system to satisfy with the diversity of the maritime environments [7], [13], [14], [27], [36]. $D_{S1}$ stands for the MARVEL dataset and the new class added based on the classified categories of the naval ship dataset. There are at least ten samples in the new class added. $D_{S2}$ is the VAIS dataset. $D_{S1}$, $D_{S2}$ constitute the source dataset $D_{S}$. The label of samples in $D_{S2}$ that unmatched in the class of $D_{S1}$ are deleted to be included in the target dataset $D_{T}$. The majority of the target dataset $D_{T}$ originates from $D_{T1}$ that is the result of a web crawler. Therefore, a new dataset $D_{N}$, the VMI dataset, is obtained from the result of the merging dataset network.

The VMI dataset consists of 3750 images, 25 fine-grained categories. Every fine-grained categories contains 150 images to resolve class imbalance and 25 fine-grained categories are Passenger ship, Sailing vessel, Sailing boat, Speed boat, Jetski, Cabin boat, Windsurfer, Yacht, Barge, Passenger and ro-ro cargo ship, Bulk carriers, Container ship, Dredging ship, Fishing vessel and boat, Open hatch cargo ship, Platform supply ship, Ro-ro cargo ship, Tanker, Tugboat, Aircraft carrier, Combat vessel, Destroyer, Replenishment vessel, Buoy and Platform, seen in Fig. 6. The dataset have equal number of samples in each class. The majority of images in the dataset is more excellent than 50,176 (224 × 224) pixels to satisfy input resolution. 225 images is less than 50176 pixels, shown in FIGURE 6.
FIGURE 6. Illustration images from each fine-grained categories in the VMI dataset, consisting of (a) Aircraft carrier, (b) Barge, (c) Bulk carriers, (d) Buoy, (e) Cabin boat, (f) Combat vessel, (g) Container ship, (h) Destroyer, (i) Dredging ship, (j) Fishing vessel and boat, (k) Jet ski, (l) Open hatch cargo ship, (m) Passenger and ro-ro cargo ship, (n) Passenger ship, (o) Platform, (p) Platform supply ship, (q) Replenishment vessel, (r) Ro-ro cargo ship, (s) Sailing boat, (t) Sailing vessel, (u) Speed boat, (v) Tanker, (w) Tugboat, (x) Windsurfer, (y) Yacht.

FIGURE 7. Distribution of image resolution in the VMI dataset which is less than 50,176 pixels.

As can be seen in Table 2, (85%, 15%) is the best average results. In case 1 and case 2, the divided datasets are used with the same proportion and different images between training and test set to test the benchmark network AlexNet. In case 1, AlexNet without deep transfer learning achieves a maximum accuracy of about 10% to demonstrate weak classification performance with severe underfitting. In case 2, AlexNet based on network-based transfer learning achieves an accuracy of 79.16% in 100 epochs. In five type, three cases (70%, 30%), (75%, 25%), (80%, 20%), (85%, 15%), (90%, 10%) are underfitting and only one case (90%, 10%) is overfitting, that the extracted feature of the training set obtain large $|U|$ to overfit.

Therefore, (85%, 15%) is chosen to divide the training and test set by random sampling, and the same maritime images only belong either to the training or test dataset but...
not both. The training set has 3188 images, and the test set has 562 images. In other words, the average class imbalance ratio between different categories is less than 0.7%. However, the merging datasets network might misclassify the target dataset $D_T$ and delete correct classified images of $D_S$ for image balance. The label noise appears in the VMI dataset. Deep networks are known to be relatively robust to label noise, which has little effect on the results [37]. Therefore, the backbone networks evaluated and analyzed the VMI dataset.

Then, the classification capabilities of the backbone networks are tested three times in the VMI dataset. The backbone networks, including AlexNet, VGGNet (VGGNet-11, VGGNet-16), ResNet (ResNet-18, ResNet-50, ResNet-101), DenseNet (DenseNet-121, DenseNet-169), SqueezeNet, and EfficientNet (EfficientNet-b0, EfficientNet-b5, EfficientNet-b7), are measured by accuracy among 100 epochs to obtain the maximum accuracy every 25 epochs. The backbone networks based on network-based transfer learning are pre-trained in ImageNet. We use Adam optimizer with 0.001 learning rate, and cross-entropy loss function, and $224 \times 224$ input size is chosen to train the backbone networks. The results are shown in Fig. 8, Table 3, and Fig. 9.

**Fig. 8.** Accuracy of the proposed backbone networks to show data dispersion during the training. The maximum, upper quartile, median, lower quartile, and minimum is shown from right to left. Outliers are represented by small circles.

Fig. 8 indicates the relationship between the accuracy and the training epochs of the backbone networks. The bar in the box diagram shows maximum, upper quartile, median, lower quartile, and minimum from right to left. The maximums represent the accuracy that the backbone networks may achieve in the final training, and the minimums mean the minimum accuracy of the backbone networks during the training. The data between the maximums and the minimums show the degree of change in the accuracy of the model during the training. The accuracy is easier to concentrate between the upper quartile and the lower quartile to indicate a greater variation in the training, such as VGGNet-16. The medians represent the relationship between the training epochs and the difficulty of the maximum accuracy, and show that the distance between the median value and the maximum value is close for less training epochs, such as ResNet-50 and SqueezeNet. The outliers, which represent training instability, are outside the range of maximum and minimum, such as ResNet-18, DenseNet-169, and SqueezeNet. The probability of the outliers is less than 0.7%. Intuitively, EfficientNet reach better performance than the other backbone networks, and the accuracy of EfficientNet-b5 is the highest. These results confirm EfficientNet transfer well and achieve state-of-the-art accuracy [22].

**Table 3.** Accuracy measures obtained for the proposed CNN approaches using the VMI dataset. The parameters (Params) and FLOPs of the backbone networks are shown.

| name       | ACC(%) | Params(M) | FLOPs(B) |
|------------|--------|-----------|----------|
| AlexNet    | 79.97  | 57        | 0.7      |
| VGGNet-11  | 85.47  | 129       | 7.6      |
| VGGNet-16  | 86.06  | 134       | 15.5     |
| ResNet-18  | 82.20  | 11        | 1.8      |
| ResNet-50  | 85.04  | 24        | 3.8      |
| ResNet-101 | 87.23  | 43        | 7.9      |
| DenseNet-121 | 87.47 | 7         | 0.6      |
| DenseNet-169 | 89.79 | 14        | 0.7      |
| SqueezeNet | 83.78  | 1         | 0.8      |
| EfficientNet-b0 | 88.89 | 4         | 0.4      |
| EfficientNet-b5 | 91.39 | 28        | 9.9      |
| EfficientNet-b7 | 90.49 | 64        | 37       |

**Fig. 9.** Ball chart reporting the accuracy versus the parameters and FLOPs of the backbone networks. The size of each ball corresponds to the FLOPs of the model.
due to the memory of GPU limited by the parameters and FLOPs. Simultaneously, under similar parameters, the improvement of the network structure achieves the higher accuracy, such as improved depth of VGGNet [20], deep residual learning framework of ResNet [23], dense block of DenseNet [24], and fire layer of SqueezeNet [17], shown in Fig. 4. Therefore, compared with the other backbone networks, the EfficientNet-b5 reaches the highest accuracy with similar parameters and FLOPs. Then, precision, recall, and F-measure are evaluated in EfficientNet-b5 to analyze the classification of each category.

The confusion matrix reveals the categories to which each image in the test set is assigned, and off-diagonal categories are wrongly classified, shown in Fig. 10. The class bulk carrier and many other classes have misclassifications with each other, such as tanker and open hatch cargo ship. The vessels are used to load bulk cargo and have a similar appearance resulting in an error of classification. At the same time, the precision, recall, and F-1 measure data indicate that different types of images have different prediction labels in Table 4. The majority of classes have high similarity in the same class. Seventeen categories have about 91%-100% of F-1 values, and the accuracy rate of EfficientNet is 91.39%. However, the F-1 measure of eight categories, consisting of barge, bulk carriers, combat vessel, destroyer, fishing vessel and boat, jetski, open hatch cargo ship, and tanker, are less than 0.91.

The backbone network EfficientNet-b5 is trained and tested in the VAIS dataset to validate the effectiveness of the proposed method. The training process of the backbone networks use the same parameters as the VMI datasets. Adam optimizer with 0.001 learning rate and cross-entropy loss function is chosen, and 224 × 224 image resolution is used to input size of the proposed method. The accuracy of a few state-of-the-art methods is compared in the VAIS dataset, shown in Table 5. In contrast to an accuracy of 88.00% achieved by the previous best method Gabor + MS-CLBP + CNN + SVM, the proposed method EfficientNet-b5 using network-based deep transfer learning obtains an accuracy of 98.53%.

The performance of the two methods, which are the previous best method and the proposed method, are analyzed by the precision, recall, F-1 measure, and execution time. Figure. 11 shows that the confusion matrix reveals the class to which each image is assigned in the test set of the VAIS dataset, and off-diagonal categories are wrongly classified. The left confusion matrix is the test result of Gabor + MS-CLBP + CNN + SVM [18]. The right confusion matrix is the test result of the proposed method EfficientNet-b5 using network-based deep transfer learning. The 11 samples are only misclassified among the 750 samples in the VAIS dataset's test set using the proposed method. Therefore, the precision, recall, and F-1 measure in each class is calculated from the confusion matrix.

Table 6 reports the precision, recall, F-1 measure, and execution time in the Gabor + MS-CLBP + CNN + SVM
The VMI dataset might be altered into five coarse-grained categories, consisting of passenger (passenger ship, sailing vessel, sailing boat, speed boat, jetski, cabin boat, windsurfer, yacht), passenger/freight (barge, passenger, and ro-ro cargo ship), merchant (bulk carriers, container ship, dredging ship, fishing vessel and boat, open hatch cargo ship, platform supply ship, ro-ro cargo ship, tanker, tugboat), navy (aircraft carrier, combat vessel, destroyer, replenishment vessel) and other (buoy, platform). However, class imbalance appears in the classification of five coarse-grained categories.

Through experimental results, the backbone network EfficientNet-b5 is not possible to classify ships with the same ship style but different cargo loads, such as bulk carriers and tankers, which dominates the sea transportation of bulk cargo [12], [27]. Besides, the maritime dataset may include label noise [37], [39]. The EO camera can only capture images within a few kilometers, and it needs to have enough light sources to collect data [40]. Also, in the case of fog, snow, and rain, the image quality of EO is blurred and not easily identified due to being limited by camera parameters [7]. Fortunately, a few methods to reduce label noise [41], [42], a small object detection method [43], increasing the brightness of night image [44], dehazing [45], rain and snow removal [46] are proposed, which have strengthened the application of EO camera for image processing and computer vision. So, the new ship classifier will be integrated with not only the image captured from the EO camera but also the information of multi-sensor data fusion to improve image classification and recognition.

**TABLE 6.** Comparison of the precision (P), recall (R), F-1 measure (F-1), and execution time (seconds) with the two methods, which are Gabor + MS-CLBP + CNN + SVM and the proposed method, in the VAIS dataset.

| Method                  | Gabor + MS-CLBP + CNN + SVM [18] | Proposed Method |
|-------------------------|----------------------------------|-----------------|
| medium "other" ship     | 84.62 51.16 63.77 97.65 96.51 97.08 | 89.73 92.94 90.73 95.23 94.75 95.91 |
| medium passenger ship   | 70.13 87.10 77.70 95.38 100.0 97.64 | 83.75 92.35 91.75 95.35 93.25 94.75 |
| merchant ship           | 92.96 92.96 92.96 97.26 100.0 98.61 | 85.75 93.25 92.75 97.15 95.25 95.75 |
| sailing ship            | 98.48 97.98 98.23 100.0 99.49 99.75 | 87.55 95.25 94.75 98.95 97.25 97.55 |
| small boat              | 86.31 92.65 89.37 100.0 97.76 98.87 | 84.55 93.15 92.65 97.85 96.25 96.75 |
| tugboats                | 70.59 60.00 64.87 83.33 100.0 90.90 | 76.55 68.25 66.75 81.95 99.25 98.75 |
| Execution time          | $36.73$ | $89.14$ |

and the proposed method. The execution time [18] indicates training and test time of the VAIS dataset to achieve the best performance, which evaluates the performance of the model mixed with other algorithms, rather than FLOPs that only reflect the performance of CNNs. Compared with Gabor + MS-CLBP + CNN + SVM, the precision, recall, F-1 measure using the proposed method are improved, achieving better performance. The execution time of the proposed method is about 892 seconds, which is similar to the previous method. The backbone network EfficientNet-b5 using network-based deep transfer learning is proved to achieve state-of-the-art accuracy, precision, recall, and F-1 measure, tested in the EO images of the VAIS dataset.

**VI. CONCLUSION AND DISCUSSION**

We proposed the double transfer method to obtain a new dataset, the Visible Maritime Image dataset, consisting of ship, vessels, boat, and others of maritime environments, to expand the scope of maritime image classification and conform to the actual situation. The accuracy of the test set obtained 91.39% using the backbone network EfficientNet-b5 by deep transfer learning to verify that small fine-grained datasets can achieve higher accuracy. Simultaneously, the backbone network EfficientNet-b5 using the same parameters was trained on the VAIS dataset to achieve the best accuracy of 98.53%, which is improved by about 6% to 10% than other algorithms. It proves that the backbone network is robust.

The VMI dataset might be altered into five coarse-grained categories, consisting of passenger (passenger ship, sailing vessel, sailing boat, speed boat, jetski, cabin boat, windsurfer, yacht), passenger/freight (barge, passenger, and ro-ro cargo ship), merchant (bulk carriers, container ship, dredging ship, fishing vessel and boat, open hatch cargo ship, platform supply ship, ro-ro cargo ship, tanker, tugboat), navy (aircraft carrier, combat vessel, destroyer, replenishment vessel) and other (buoy, platform). However, class imbalance appears in the classification of five coarse-grained categories.

Through experimental results, the backbone network EfficientNet-b5 is not possible to classify ships with the same ship style but different cargo loads, such as bulk carriers and tankers, which dominates the sea transportation of bulk cargo [12], [27]. Besides, the maritime dataset may include label noise [37], [39]. The EO camera can only capture images within a few kilometers, and it needs to have enough light sources to collect data [40]. Also, in the case of fog, snow, and rain, the image quality of EO is blurred and not easily identified due to being limited by camera parameters [7]. Fortunately, a few methods to reduce label noise [41], [42], a small object detection method [43], increasing the brightness of night image [44], dehazing [45], rain and snow removal [46] are proposed, which have strengthened the application of EO camera for image processing and computer vision. So, the new ship classifier will be integrated with not only the image captured from the EO camera but also the information of multi-sensor data fusion to improve image classification and recognition.

**REFERENCES**

[1] Q. Du, Y. Zhang, X. Yang, and W. Liu, “Ship target classification based on Hu invariant moments and ART for maritime video surveillance,” in Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS), BattJ, AB, Canada, Aug. 2017, pp. 414–419.

[2] A.-J. Gallego, A. Pertusa, and P. Gil, “Automatic ship classification from optical aerial images with convolutional neural networks,” Remote Sens., vol. 10, no. 4, p. 511, Mar. 2018.

[3] M. Leclerc, R. Thamarasa, M. C. Florea, A.-C. Boury-Brisset, T. Kirubarajan, and N. Duclos-Hindie, “Ship classification using deep learning techniques for maritime target tracking,” in Proc. 21st Int. Conf. Inf. Fusion (FUSION), Cambridge, U.K., Jul. 2018, pp. 737–744.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., Vancouver, BC, Canada, Dec. 2012, pp. 1097–1105.

[5] Y. Wang, Q. Yao, J. Kwok, and L. M. Ni, “Generalizing from a few examples: A survey on few-shot learning,” 2019, arXiv:1904.05046. [Online]. Available: http://arxiv.org/abs/1904.05046

[6] S. Moosbauer, D. Konig, J. Jakel, and M. Teutsch, “A benchmark for deep learning based object detection in maritime environments,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Long Beach, CA, USA, Jun. 2019, pp. 916–925.

[7] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, “Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 8, pp. 1993–2016, Aug. 2017.

[8] L. Patino, T. Cane, A. Vallee, and J. Ferryman, “PETS 2016: Dataset and challenge,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Las Vegas, NV, USA, Jun. 2016, pp. 1240–1247.

[9] M. Kristan, V. Sulek Kenk, S. Kovacic, and J. Pers, “Fast image-based obstacle detection from unmanned surface vehicles,” IEEE Trans. Cybern., vol. 46, no. 3, pp. 641–654, Mar. 2016.
[10] B. Bovcon, R. Mandelj, J. Perš, and M. Kristan, “Stereoscopic obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation,” Robot. Auto. Syst., vol. 104, pp. 1–13, Jun. 2018.

[11] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, “ARGOS-venice boat classification,” in Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveillance (CVPRW), Boston, MA, USA, June 2015, pp. 10–16.

[12] C. Dao-Duc, H. Xiao-bai, and O. Morére, “Maritime vessel images classification using deep convolutional neural networks,” in Proc. 6th Int. Symp. Inf. Commun. Technol. (SoICT), New York, NY, USA, 2015, pp. 276–281.

[13] M. M. Zhang, J. Choi, K. Danilidis, 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 (CVPRW), Boston, MA, USA, Jun. 2015, pp. 1–6.

[14] E. Gundogdu, B. Solmaz, V. Yücesoy, and A. Koc, “Marvel: A large-scale image dataset for maritime vessels,” in Prof. Asian Conf. Comput. Vis. (ACCV), Taipei, Taiwan, Nov. 2016, pp. 165–180.

[15] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” Neural Netw., vol. 106, pp. 249–259, Oct. 2018.

[16] 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., Miami, FL, USA, Jun. 2009, pp. 248–255.

[17] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and<0.5mb model size,” in Proc. Int. Conf. Learn. Represent. (ICLR), Long Beach, CA, USA, Apr. 2017, pp. 1–13.

[18] Q. Shi, W. Li, F. Zhang, W. Hu, X. Sun, and L. Gao, “Deep CNN with multi-scale rotation invariance features for ship classification,” IEEE Access, vol. 6, pp. 38656–38668, 2018.

[19] Q. Shi, W. Li, R. Tao, X. Sun, and L. Gao, “Ship classification based on multifeature ensemble with convolutional neural network,” Remote Sens., vol. 11, no. 4, pp. 419–430, Feb. 2019.

[20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, May 2015, pp. 1–14.

[21] M. Milicevic, K. Zubrinic, I. Obradovic, and T. Sjekavica, “Data augmentation and transfer learning for limited dataset ship classification,” in WSEAS Trans. Syst. Control., vol. 13, no. 1, p. 460, 2018.

[22] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn., Long Beach, CA, USA, May 2019, pp. 6105–6114.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[24] H. Zhuang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

[25] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis for unmanned surface vehicles by IMU-assisted semantic segmentation,” in Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveillance (CVPRW), Boston, MA, USA, June 2015, pp. 1134–1142.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. 6th Int. Symp. Inf. Commun. Technol. (SoICT), New York, NY, USA, 2015, pp. 276–281.

[27] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent. (ICLR), Long Beach, CA, USA, Apr. 2017, pp. 1–13.

[28] K. Rainey, J. D. Reeder, and A. G. Corelli, “Convolution neural networks for ship type recognition,” in Proc. 26th Autom. Target Recogn., Baltimore, MD, USA, vol. 9844, May 2016, Art. no. 984409.

[29] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, “Deep joint rain detection and removal from a single image,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 1685–1694.

[30] K. Rainey, J. D. Reeder, and A. G. Corelli, “Convolution neural networks for ship type recognition,” in Proc. 26th Autom. Target Recogn., Baltimore, MD, USA, vol. 9844, May 2016, Art. no. 984409.

[31] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, “Deep joint rain detection and removal from a single image,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 1685–1694.

[32] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, May 2015, pp. 1–15.

[33] R. Zhao, Y. Yang, Y. Cao, X. Li, A. C. Kot, and J. Yang, “Deep残差网络：一种新颖的残差网络结构设计。” in Proc. Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, May 2015, pp. 1–15.
JIANHUA WANG received the B.E. degree in fluid control and manipulation from Beihang University, Beijing, China, in 1987, the M.S. degree in pattern recognition and intelligent system from Chongqing University, Chongqing, China, in 2004, and the Ph.D. degree in pattern recognition and intelligent system from Shanghai Jiao Tong University, Shanghai, China, in 2007.

Since 2007, he has been an Associate Professor with the Institute of Logistics Science and Control Engineering, Shanghai Maritime University, Shanghai. He is the author more than 30 articles and holds more than 15 patents. His research interests include navigation, guidance, and control for unmanned surface vehicles, consisting of object detection and tracking, SLAM, path planning and tracking, and embedded systems.

XIANG ZHENG received the B.E. degree in automation from Beihang University, Beijing, China, in 2003, the M.S. degree in advanced control and system engineering from UMIST, Manchester, U.K., in 2004, and the Ph.D. degree in control engineering from Loughborough University, Loughborough, U.K., in 2011.

Since 2014, she has been a Lecturer with the Institute of Logistics Science and Control Engineering, Shanghai Maritime University, Shanghai, China. Her current research interest includes applications of advanced control for unmanned surface vehicles.

JING WEN received the B.E. degree in electrical engineering from Shanxi University, Taiyuan, China, in 2018. He is currently pursuing the M.S. degree in detection technology and automatic equipment with Shanghai Maritime University, Shanghai, China.

LIUZHONG RAO received the B.E. degree in measurement and control technology and instrumentation specialty from Weifang University, Weifang, China, in 2018. He is currently pursuing the M.S. degree in detection technology and automatic equipment with Shanghai Maritime University, Shanghai, China.

JUNXIA ZHAO received the B.E. degree in automation from Changzhou University, Changzhou, China, in 2018. She is currently pursuing the M.S. degree in detection technology and automatic equipment with Shanghai Maritime University, Shanghai, China.