

A Deep Learning Model to Extract Ship Size From Sentinel-1 SAR Images

Yibin Ren¹, Member, IEEE, Xiaofeng Li², Fellow, IEEE, and Huan Xu³, Student Member, IEEE

Abstract—This study develops a deep learning (DL) model to extract the ship size from Sentinel-1 synthetic aperture radar (SAR) images, named SSENet. We employ a single shot multibox detector (SSD)-based model to generate a rotatable bounding box (RBB) for the ship. We design a deep-neural-network (DNN)-based regression model to estimate the accurate ship size. The hybrid inputs to the DNN-based model include the initial ship size and orientation angle obtained from the RBB and the abstracted features extracted from the input SAR image. We design a custom loss function named mean scaled square error (MSSE) to optimize the DNN-based model. The DNN-based model is concatenated with the SSD-based model to form the integrated SSENet. We employ a subset of the OpenSARShip, a data set dedicated to Sentinel-1 ship interpretation, to train and test SSENet. The training/testing data set includes 1500/390 ship samples. Experiments show that SSENet is capable of extracting the ship size from SAR images end to end. The mean absolute errors (MAEs) are under 0.8 pixels, and their length and width are 7.88 and 2.23 m, respectively. The hybrid input significantly improves the model performance. The MSSE reduces the MAE of length by nearly 1 m and increases the MAE of width by 0.03m compared to the mean square error (MSE) loss function. Compared with the well-performed gradient boosting regression (GBR) model, SSENet reduces the MAE of length by nearly 2 m (18.68%) and that of width by 0.06 m (2.51%). SSENet shows robustness on different training/testing sets.

Index Terms—Custom loss function, deep learning (DL), deep neural network (DNN) regression, ship size extraction, synthetic aperture radar (SAR) image.

I. INTRODUCTION

SHIP detection is of great significance to marine activities, such as marine transportation, fishery management, and maritime safety [1]. Spaceborne Synthetic Aperture Radar (SAR) can monitor targets under all-day and all-weather

Manuscript received October 2, 2020; revised November 30, 2020 and January 14, 2021; accepted February 26, 2021. Date of publication March 17, 2021; date of current version December 9, 2021. This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19060101, Grant XDB42040401, and Grant XDA19090103; in part by the Key Research and Development Project of Shandong Province under Grant 2019JZZY010102; in part by the Key Deployment Project of Center for Ocean Mega-Science; in part by the Chinese Academy of Sciences (CAS) under Grant COMS2019R02; in part by the CAS Program under Grant Y9KY04101L; and in part by the China Postdoctoral Science Foundation under Grant 2019M662452. (Corresponding author: Xiaofeng Li.)

Yibin Ren and Xiaofeng Li are with the Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China, and also with the Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China (e-mail: yibinren@qdio.ac.cn; xiaofeng.li@ieee.org).

Huan Xu is with the School of Geomatics and Marine Information, Jiangsu Ocean University, Lianyungang 222005, China (e-mail: 2018224050@jou.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3063216

conditions, making it one of the most critical marine surveillance tools [2]. Ship detection from SAR images has always been a hotspot in marine applications [3]–[6]. With the advent of a new generation of satellites and fast-growing image analysis technology, it is feasible to extract more detailed ship information from SAR images in addition to its detection [7]. The ship's length and width provide essential information for ship classification and marine surveillance [8]. In most cases, it is difficult to identify the ship's type directly from the SAR image, and the size information can provide useful help. In addition, the elaborate geometric parameter estimation is also meaningful to SAR imagery interpretation. With the rapid increase in SAR data volume, an efficient and accurate ship size extraction method will bring a new idea for SAR image interpretation.

Generally, as metallic objects, ships can reflect the electromagnetic waves of SAR sensors much more strongly than the surrounding water. Therefore, one can perceive most ships on SAR images as bright backscattering intensity targets, characterized by the highly normalized radar cross section (NRCS) values. The geometric feature of the ship's NRCS, such as the length and width of the minimum bounding rectangle (MBR), provides an initial size for estimating a ship's ground size. Meanwhile, the ship's superstructure, the environment, the sea–ship interaction, and the imaging system can influence the NRCS from the ship [8]. These factors lead to a huge gap between the initial size and the ground size. Therefore, it is challenging to extract the ship size from SAR images accurately.

In the literature, all SAR ship size detection algorithms include three steps: 1) binarization, 2) initial size extraction, and 3) accurate size estimation. The first procedure processes the SAR image and divides the pixels into ship signatures and nonship signatures. The second procedure extracts the initial ship size based on the binary results, such as the MBR of the ship signature. The third step estimates the accurate size based on the initial ship size and other relevant factors. Stasolla and Greidanus [9] employed the constant false alarm rate (CFAR) to do the first step. The CFAR family is [10]–[12] widely used in ship detection on SAR images to separate ship signature and background. Furthermore, for the second step, they employed the mathematical morphology method to refine the signature to extract the MBR of the ship. They did not develop the third step and adopted the MBR's length and width as the vessel's estimated length and width. The method was tested with 127 available ship samples from Sentinel-1 SAR images. The mean absolute error (MAE) of length measurement is 30 m (relative error: 16%), and the MAE of width is 11 m

(relative error: 37%). To further reduce the estimation error, researchers took the initial ship size of the second step as the initial information and performed the third step by statistical/machine learning (ML) methods, such as multiple linear regression [13], the kernel-based method [7], and nonlinear regression [8]. Among them, the nonlinear regression model is a typical representation. In 2018, Li *et al.* [8] used OpenSARShip [14], a large-volume data set dedicated to Sentinel-1 ship interpretation, to extract the ship's length and width. First, they performed the binarization step by a threshold-based method and obtained the ship signature. Second, they refined the ship signature by an image segmentation operation and obtained the initial ship size. Third, they fused the dual-polarization information and considered the initial ship size, the environment information, sensor information, etc. to construct a nonlinear regression model based on the gradient boosting. The gradient boosting is an ensemble learning method, allowing for the optimization of arbitrary differentiable loss functions [15]. The nonlinear regression model established an accurate mapping between the input factors and the ship length and width. They validated their results against ships longer than 80 m. Experiments showed that the MAEs of length and width are 8.80 m (relative error: 4.66%) and 2.17 m (relative error: 7.01%), respectively. As the pixel spacing of the Sentinel-1 images in their database is 10 m, the model made noticeable progress and pushed the error under one-pixel spacing.

Overall, the accuracy of the extracted size is improved continuously. With the rapid increase in satellite data, the traditional three-step procedure shows an obvious limitation: it is extremely complex. Each step's error may accumulate to the final extracted size. In most cases, the first two steps, binarization and initial size extraction, need sophisticated image operations to meet the following estimation step [8]. For the third step, selecting influence factors to construct conventional ML-based regression models requires high-level prior expert knowledge and manual engineering [16], which is also a big challenge. The errors generated in each step will be accumulated and ultimately affect the final size extraction accuracy. Therefore, in the era of big data, it is feasible to develop new methods to improve ship size extraction accuracy and efficiency.

ML has gradually evolved into deep learning (DL) that brings new ideas for addressing the above challenge [17]. A typical DL model consists of deep neural networks (DNNs), accepts input data in a raw format, and automatically learns the required features to achieve classification or prediction [18]. This process is known as end-to-end learning. Compared with ML, DL significantly simplifies feature engineering and is suitable for modeling big data and complex relationships. In recent years, DL has been successfully applied in oceanography, geography, and remote sensing, which has helped people gain further process understanding of Earth system science problems [19]–[25].

A deep convolution neural network (CNN) is a particular type of DNN composed of CNN layers. A CNN layer connects to the previous layer's local patches through convolution kernels to extract local spatial features [26]. Since deep

CNN models have achieved great success in image classification [27], researchers proposed various ship detection models based on typical deep CNN frameworks. Representations include a faster region-based convolutional network (Faster-RCNN) [28], [29], a single shot multibox detector (SSD) [30], [31], you only look once (YOLO) [32], and other integrated detection frameworks [33], [34]. These DL-based models achieved good performance in detecting ships from SAR images, with an average precision (AP) over 80%, which is an obvious improvement than the classical models. To further improve the performance, researchers introduced a rotatable bounding box (RBB) to replace the traditional nonrotating RBB in the DL detection frameworks. Liu *et al.* optimized the traditional SSD [35] by rotating the prior box [36] and developed DRBox. Compared with the Faster-RCNN and the traditional SSD, DRBox improves the AP by 11.77% and 11.17% in detecting densely arranged ships, respectively. Furthermore, DRBox-v2 [37] optimized DRBox by integrating the feature pyramid network (FPN) [38] and focal loss (FL) [39]. Validation experiment based on the SAR ship detection data set (SSDD) showed that the AP of DRBox-v2 is 9.36% and 6.4% higher than the rotation dense feature pyramid networks (R-DFPNs) [40] and the DRBox, reaching 92.81%.

DL has become a mainstream solution for a typical ship detection task. Compared with conventional models, DL-based models significantly simplify feature engineering and achieve end-to-end detections with higher accuracy and better robustness. Inspired by this, we construct an end-to-end DL model to replace the traditional three-step procedure for ship size extraction from SAR images. There are two contributions to our study. First, we develop an end-to-end DL-based model, SSENNet, to extract the ship size from SAR images. SSENNet employs DRBox-v2 to generate the ship's RBB from the SAR image and constructs a DNN-based regression model to estimate the accurate ship size. To the best of our knowledge, SSENNet is the first end-to-end model for the ship size extraction of satellite images. Second, we design a hybrid input and a loss function named mean scaled square error (MSSE) for the DNN-based regression model, which significantly improves ship size estimation accuracy.

The rest of this article is organized as follows. Section II describes the data. Section III presents SSENNet in detail. In Section IV, experiments are conducted to evaluate the effectiveness of SSENNet. Section V discusses the performances of different regression models. The factors that influence the model performances are also discussed. Finally, Section VI concludes this article.

II. DATA

A. OpenSARShip Database

The experimental data is a subset of the OpenSARShip. The OpenSARShip (<http://opensar.sjtu.edu.cn/>) is a data set dedicated to Sentinel-1 ship interpretation, providing 11 346 SAR ship chips integrated with the automatic identification system (AIS) messages. The Sentinel-1 images are the ground range detected (GRD) products of an IW mode. The spatial resolution is about 20 m, and the pixel spacing is 10 m.

The preprocessing procedures, such as radiometric calibration and terrain correction, are carried out by SNAP 3.0. Each SAR chip contains one ship and is stored in a matrix, which indicates the amplitude values of pixels for VH (vertical emitting and horizontal receiving) and VV (vertical emitting and vertical receiving) polarizations. Here, we use 1890 samples in the VV mode as the experimental data set. Fig. 1(a), (d), (g), and (j) shows four types (cargo, tanker, tug, and fishing) of ships as examples. The AIS messages provide the ground size for each ship, which has been integrated into the OpenSARShip and can be obtained directly. Fig. 1(c), (f), (i), and (l) are the ground size corresponding to Fig. 1(a), (d), (g), and (j). Fig. 1(m) and (n) shows the distributions of length and width. The minimum and maximum values of length are 28 and 399 m, respectively. For width, the minimum and maximum values are 6 and 65 m, respectively. To match the requirement for the input size of DRBox-v2, we expand the size of each ship chip to 300×300 pixels. The SAR image values are scaled to $[0, 255]$.

B. Labeling

The ground truths for training and testing SSENet include two parts. The first one is the ground ship size, which can be obtained directly from the OpenSARShip. The other one is the ground RBB for each ship. An RBB is a rotated bounding box surrounding the ship's signature on the SAR image. The DRBox-v2 is the first part of SSENet, which generates the RBB of each ship. The ground RBB is used to train the DRBox-v2 to generate an accurate RBB. As the OpenSARShip has no RBB, we manually label the RBB for each ship by a MATLAB tool shared in DRBox-v2 [37]. The rule for labeling an RBB is to surround the ship's signature on the SAR image as precisely as possible. Fig. 1(b), (e), (h), and (k) shows the labeled RBB of each ship. The angle range of labeled RBBs is $(0^\circ, 180^\circ)$. It is worth noting that there is an apparent gap between the size of the labeled RBB and the ship's ground size [Fig. 1(b)/(c), (e)/(f), (h), and (k)]. Therefore, as stated earlier, based on ship detection, ship size extraction still needs a lot of challenging efforts.

III. METHOD

The general structure of SSENet includes three steps (see Fig. 2): 1) generating RBBs, 2) estimating the ship size based on a DNN-based regression model, and 3) calculating MSSE loss and optimizing SSENet. The first step takes the SAR chip as input and automatically detects the ship's RBBs by a deep CNN model, DRBox-V2. Then, the RBB with the highest confidence is selected as the initial RBB. The second step estimates the ship size by a DNN model. We construct a hybrid input for the DNN model. The hybrid input consists of two parts: 1) the initial ship parameters (the ship size and orientation angle) obtained from the initial RBB and 2) the feature map extracted in the first step. The DNN model then predicted width and length of the ship. The third step calculates the MSSE loss of the estimated size. The MSSE loss is summed with confidence, and the location losses are calculated in the first step to form the final loss. All trainable

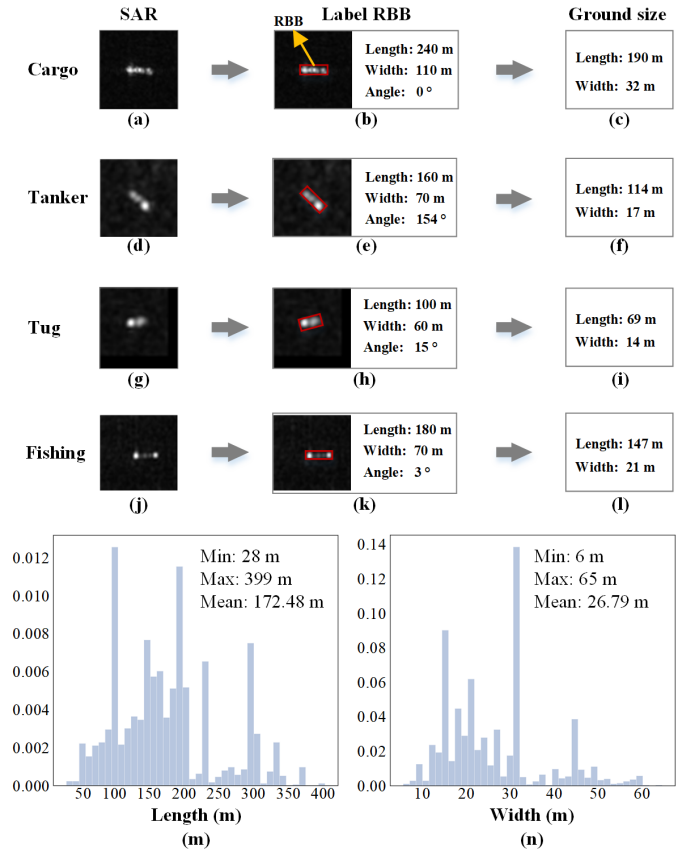


Fig. 1. Examples of SAR image chips of typical ships (cargo, tanker, tug, and fishing) and their corresponding labeled RBB and ground size: (a), (d), (g), and (j) SAR chips; (b), (e), (h), and (k) labeled RBBs and their sizes; (c), (f), (i), and (l) ground sizes of the ships. (m) The length distribution of the 1890 experimental samples. (n) The width distribution of the 1890 experimental samples.

parameters of SSENet are optimized by minimizing the final loss.

Overall, the idea of SSENet is consistent with the traditional three-step ship size extraction method: obtaining the initial ship size first (first two steps) and then estimating the accurate ship size (the third step). Since the CNN has advantages in image target detection and the DNN has advantages in nonlinear regression, SSENet combines the two to achieve the size extraction in an end-to-end way.

A. Generating RBBs

Generating RBBs for the ship is based on DRBox-v2 [37]. DRBox-v2 is a ship detection model based on the SSD. It optimizes the traditional SSD by integrating the angle of ship orientation into the SSD. It outputs the RBBs of ships. The input is a 300×300 pixel SAR image. We employed VGG16, which consists of five feature extraction modules, to extract feature maps from the input image. The first feature extraction module comprises two CNN layers, and the others are composed of a max-pooling layer and two CNN layers. Five feature maps named F_1, F_2, \dots, F_5 are generated. The number of channels in the F_1-F_5 feature maps is 64, 128, 256, 512, and 512. As the kernel of four max-pooling layers

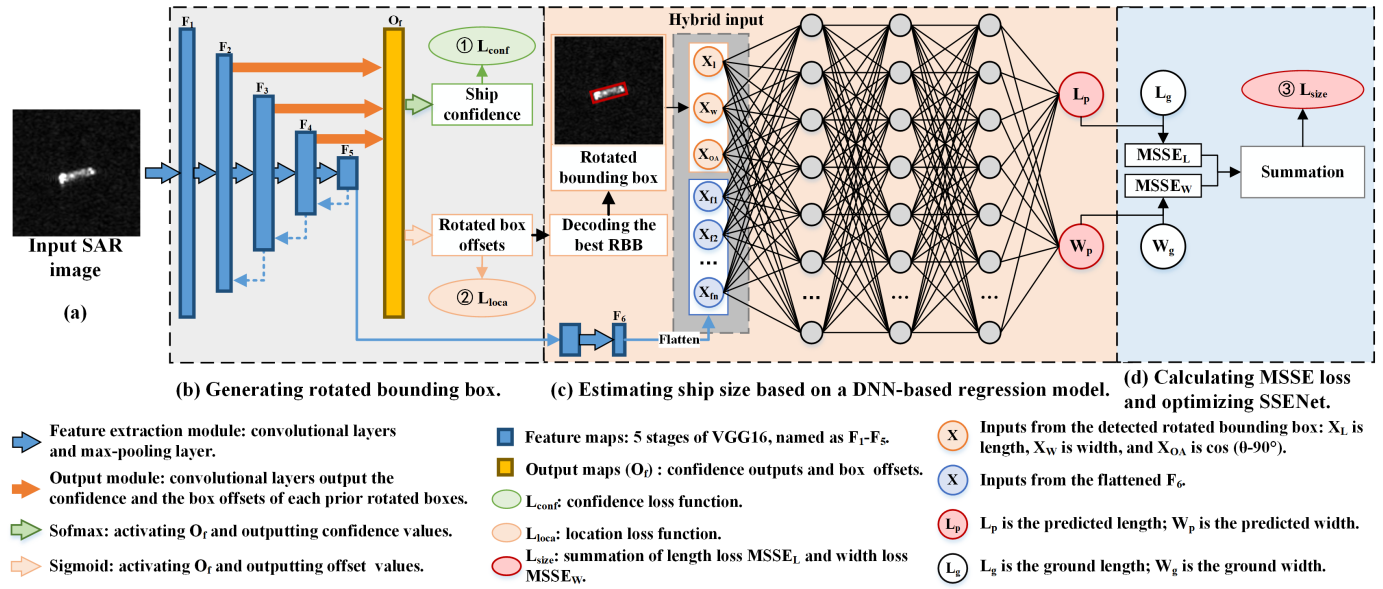


Fig. 2. Detailed structure of SSENet. (a) Input SAR image. (b) Generating RBBs. (c) Estimating ship size based on a DNN model. (d) MSSE loss function.

is 2×2 , the spatial size of F_1 - F_5 feature maps is 300×300 , 150×150 , 75×75 , 38×38 , and 19×19 pixels, respectively. The output modules perform convolution on feature maps and generate output maps O_f [Fig. 2(b)]. Softmax and sigmoid functions [41] activate O_f to create the confidence to be a ship and the location offsets of each prior RBB. We use three feature maps (F_2 - F_4) to generate O_f . FPN is employed to fuse feature maps at different levels. The cross-entropy loss and the smooth L1 loss [35] are used as the confidence loss and location loss for DRBox-v2.

After the first step, a series of candidate RBBs with confidence are obtained for a ship, which provides initial references for the subsequent accurate size estimation.

B. Estimating Ship Size Based on a DNN Model

First, we construct a hybrid input for the DNN model. The first part of the hybrid input is the initial ship size and orientation angle obtained from the best RBB, which provides primary and direct information for accurate ship size regression. In Fig. 2(c), we select the RBB with the highest confidence value from the candidate RBBs as the best RBB. The length and width of the best RBB are employed as the initial ship size. The orientation angle of the best RBB is employed to calculate the ship orientation angle. The orientation angle measures the ship orientation to the azimuth direction, as shown in Fig. 3. It affects the ship signature on the SAR image [8], [14]. It does not distinguish between the bow and the stern of one ship, and its range is $(-90^\circ, 90^\circ)$.

The other part of the hybrid input is the feature map extracted from the input SAR image. The SAR image contains the ship's signature and the sea clutter under typical environmental conditions, which provides potential information for ship size estimation. The ship's signature on the SAR image reflects the ship's state, moving or stationary. The moving target point is often located in more than one resolution

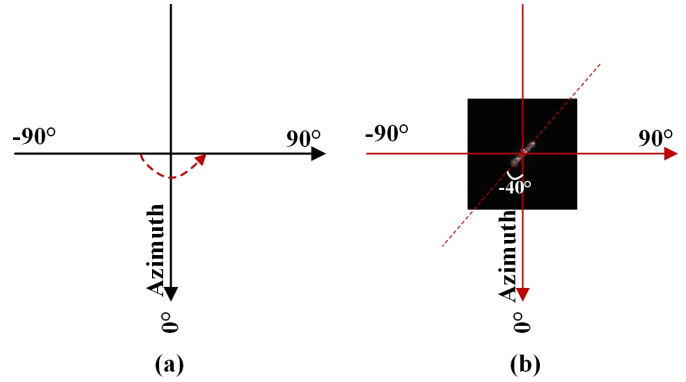


Fig. 3. Illustration of the ship orientation. (a) Coordinate system. (b) Example of a ship chip.

cell during the SAR integration time. The dispersion of the backscattered energy causes the smearing and the degradation of brightness in the SAR image. In addition, the signature of a moving ship shows an azimuth displacement. The SAR system receives the Doppler signal from the scatterer in the azimuth direction. The azimuth position of a stationary ship is consistent with the azimuth position of the SAR platform. However, there is an additional component to the Doppler shift for a moving ship, resulting in an azimuthal displacement in the ship signature. The environmental conditions during satellite imaging, such as wind fronts, ocean waves, and rain cells, affect the ship's signature on the SAR image. Under the typical conditions, the sea-ship interaction brings out a complex ship motion in the real world and relatively different polarimetric scattering mechanisms in the SAR signature. Several studies analyzed the impacts of complex environmental conditions on the ship's signature [42]-[44]. The correlation between the ship's state, the environmental conditions, and the ship's size has been confirmed in reference [8]. The abstracted feature

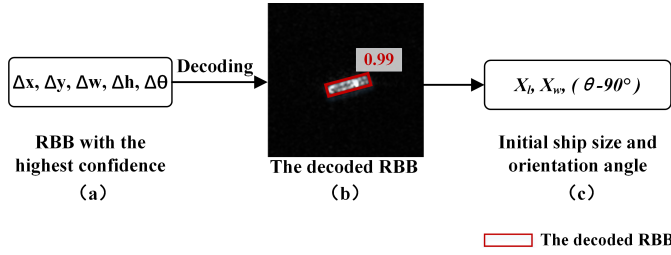


Fig. 4. Obtaining the initial ship size and orientation angle. (a) Five parameters of the best RBB. (b) Decoded RBB. (c) Initial ship size and orientation angle.

map extracted from the input SAR image contains the above factors. Therefore, we transform the high-level feature map F_5 in Fig. 2(b) as the other part of the hybrid input.

Second, we construct a DNN model to regress the accurate ship size. The DNN model is composed of several fully connected neural network (NN) layers. The hybrid input is fed into the DNN model. After layer-by-layer transformation, the DNN model output the ship's accurate length and width. In the following, we will detail the hybrid input and the DNN regression model.

1) *Constructing a Hybrid Input for the DNN Model:* Constructing a hybrid input includes two procedures: 1) obtaining the initial ship size and orientation angle and 2) transforming feature map F_5 as inputs.

1) Obtaining the initial ship size and orientation angle.

A series of candidate RBBs is generated by the DRBox-v2. The RBB with the largest confidence value is selected as the best RBB for the ship. The best RBB output by the location regression of the DRBox-v2 (SSD) is an encoded RBB, not a standard RBB [35], [37]. A standard RBB consists of five parameters: center coordinates (x, y) , length (l) , width (w) , and orientation angle (θ) . An encoded RBB consists of the offsets of these five parameters: $\Delta x, \Delta y, \Delta l, \Delta w$, and $\Delta \theta$ [Fig. 4(a)]. To obtain the standard ship size and orientation angle, we decode the encoded RBB to a standard one by the decoding transformation [37] [Fig. 4(b)]. The length (X_L) and width (X_W) of the decoded RBB will be input to the DNN regression model. As the range of the RBB angle is $(0^\circ, 180^\circ)$, we add -90° to the angle to obtain the initial ship orientation $(\theta - 90^\circ)$ [Fig. 4(c)]. Since the effect of orientation is symmetric to the azimuth, we input the cosine transformation of the ship orientation $\cos(\theta - 90^\circ)$ into the DNN regression model, X_{OA} in Fig. 2(c).

2) Transforming feature map F_5 as inputs. We transform F_5 as inputs for the DNN regression model. As stated in Section iii-A, F_5 contains 512 channels with a spatial size of 19×19 pixels. If we flatten F_5 into a 1-D input vector and directly input this vector into the DNN regression model, the input layer of the DNN model contains 184832 ($512 \times 19 \times 19$) elements. So, many input elements will increase the complexity and training difficulty of the fully connected DNN regression model.

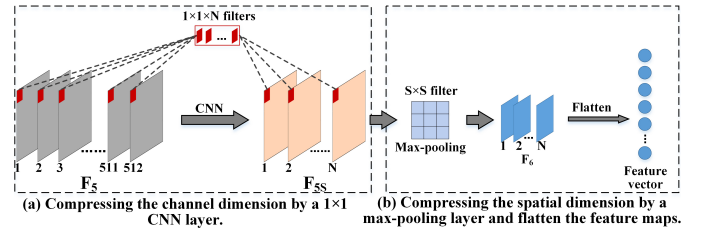


Fig. 5. Transforming feature map F_5 as inputs. (a) Compressing F_5 in the channel dimension. (b) Compressing F_5 in the spatial dimension.

To reduce the dimension of the DNN's input vector, we compress F_5 in the channel and spatial dimensions. First, we transform F_5 by a CNN layer whose convolutional kernel size is 1×1 , and the filter number is N , obtaining F_{5S} . The channel number of F_5 is reduced from 512 to N [Fig. 5(a)]. Second, we perform an S size max-pooling operation on the new F_{5S} and obtain a new feature map F_6 [Fig. 5(b)]. The spatial size of F_6 is $\lceil 19/S \rceil$. We will discuss the N and S values in the experimental section. Finally, we flatten F_6 as a 1-D feature vector, concatenated with the initial width, length, and orientation to form the DNN regression model's hybrid input.

2) *Regressing Accurate Ship Size Based on a DNN Model:*

The feature vector of F_6 is concatenated with the size vector to form the hybrid input for the DNN model [Fig. 2(c)]. Here, the DNN model has three hidden NN layers, each containing 256 neurons. The number of hidden NN layers and the number of neurons are obtained by the parameter-tuning experiment, which will be detailed in Section IV-F. The activation function of each layer is the rectified linear unit (ReLU). An output layer includes two neurons stacked on the last hidden layer. As we scaled the ship size values to 0-1, we employ a sigmoid function to activate the output layer. Finally, the predicted width W_p and the estimated length L_p are obtained [Fig. 2(c)].

C. Calculating MSSE Loss and Optimizing SSNet

We design a new loss function, MSSE, as the loss function of the DNN-based ship size regression model. For regression tasks, mean square error (MSE) is a widely adopted loss function. In (1), y_i represents the ground truth, y'_i represents the prediction value, and N means the number of values to be predicted. It is easy to see that there is no correlation between the loss value calculated by MSE and the ground truth value. For example, a ship's true length and width are 100 and 50 m, respectively. The corresponding predicted length and width are 80 and 30 m. The MSE values for length and width are the same, both at 400. They contribute equally to optimize the model parameters. However, the length of a ship is usually longer than its width and, therefore, draws more concerns than width. To improve length estimation accuracy, we hope that the length loss contributes more to optimize the model weights than the width loss.

We propose the MSSE loss function. Unlike MSE, MSSE integrates the ground truth values of ship length and width into the classical MSE. The ground truth is used as a dynamic

parameter to scale the square error. The definition of MSSE is shown in (2); y_i , y'_i , and N are the same as the definitions in MSE. For the example mentioned in the previous paragraph, the MSSE loss of length and width are 40000 and 20000, respectively. The length loss is much greater than the width loss, and the penalty on the model will be increased in the training process. Therefore, the optimization procedure is promoted,

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2 \quad (1)$$

$$\text{MSSE} = \frac{1}{N} \sum_{i=1}^N y_i \cdot (y_i - y'_i)^2 \quad (2)$$

$$L_{\text{size}} = \text{MSSE}_L + \text{MSSE}_W. \quad (3)$$

Replacing y_i and y'_i in (2) by the ground length and the predicted length of the i_{th} ship sample, we calculate the loss of length MSSE_L . Similarly, replacing y_i and y'_i by the ground width and the predicted width, we calculate the loss of width MSSE_W . The size loss (L_{size}) is the summation of MSSE_L and MSSE_W (3).

In addition to L_{size} , the confidence loss (L_{conf}) and the location loss (L_{loca}) are another two losses calculated in the first step of detecting the ship's RBB, Fig. 2(b). L_{conf} is the cross-entropy loss, and L_{loca} is the smooth L1 loss [37], [45]. These two losses are defined as follows:

$$L_{\text{conf}} = \sum_{i=1}^N c_i \log c'_i + (1 - c_i) \log (1 - c'_i) \quad (4)$$

$$L_{\text{loca}} = \frac{1}{N} \sum_i \text{smooth}_{L1}(x_i) = \begin{cases} 0.5x_i^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (5)$$

where N is the number of predicted targets, c_i is the ground confidence of a sample, c'_i is the predicted confidence of a sample, and x_i is the element-wise difference between the ground RBB and the predicted RBB. In the training procedure, the three losses, L_{size} , L_{conf} , and L_{loca} , are added to form the final loss that optimizes SSENNet integrally.

IV. EXPERIMENTS

A. Experimental Setting

The experimental data is a subset of the OpenSARShip. We randomly selected 1890 SAR chips from the OpenSARShip. Each SAR chip has one ship. We manually label the RBB of each ship, as described in Section II. The corresponding ground length and width of each ship are collected from the AIS information in the OpenSARShip. The values of length and width are scaled to 0-1. We randomly choose 1500 SAR chips as the training set. The remaining 390 chips are the testing set. The chip size is 300×300 pixels.

The model runs on a workstation with one GeForce RTX 2070 8 GB GPU. The model is coded in Python 3.6 with the TensorFlow as the DL packages. The training batch size is 6. The initial learning rate is 0.0002. During the training process, the learning rate decreases by half every 5000 training epochs. When $L_{\text{size}} < 0.001$, then $L_{\text{loca}} < 0.005$, and the composite loss < 0.01 , the training procedure stops. The

memory limitation of the GPU determines the batch size. Six is the maximum value for the 8 GB GPU memory. The initial learning rate and the training stop condition are set based on experience and fine-tuned according to the training set's loss curve.

B. Evaluation Metrics

We employ the typical absolute and relative error metrics to evaluate the model performance: MAE and mean absolute percentage error (MAPE). Assuming y_i is the ground truth, y'_i is the prediction value, and N is the number of values to be predicted, the definitions of MAE and MAPE are as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - y'_i| \quad (6)$$

$$\text{MAPE}(\%) = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - y'_i}{y_i} \right|. \quad (7)$$

C. Model Performance Test

As parameter tuning is an essential part of the DL models, we tune the hyperparameters of SSENNet and pick up a well-trained model. We will discuss the tuning procedure in Section IV-F. We use the testing set to evaluate the performance of the well-trained SSENNet. SSENNet outputs the scaled lengths and widths of all 390 testing ships. We rescale the outputs to normal values and calculate the metrics.

The results of SSENNet are shown in Fig. 6(a) and (c). The MAE of length is 7.88 m and that of width is 2.23 m. As the pixel spacing is 10 m, the MAEs of the estimated length and width are less than 0.8-pixel spacing. The MAPEs of estimated length and width are 5.53% and 8.93%, respectively. As shown in Fig. 6(a) and (c), the R^2 score values of the estimated length and width are 0.9773 and 0.9093, respectively. These high values demonstrate that the estimated ship length/width is highly consistent with the ground length/width. The R^2 score of width being smaller than that of length showing that the performance of width estimation is worse than that of length. There are two reasons; first, the ship's width is much smaller than its length. The ship's signature on the SAR image is more ambiguous in width than in length [8]. This phenomenon causes random errors in the width of the labeled RBB, affecting the accuracy of the initial width obtained from the first step of SSENNet. Second, the MSSE loss function drives the model to pay more attention to length than to width during model training, making the model fit width less than length.

To demonstrate the necessity of the DNN-based regression model, we draw the relationship between the generated RBB's size and the ship's ground size and the relationship between the labeled RBB's size and the ship's ground size, as shown in Fig. 6(b)/(e) and (c)/(f). The generated RBB is automatically detected from the SAR images in the first stage of SSENNet. The labeled RBB is manually labeled for each ship based on the principle presented in Section II-B. As shown in Fig. 6(b)/(e) and (c)/(f), the MAE of length is nearly 40 m and

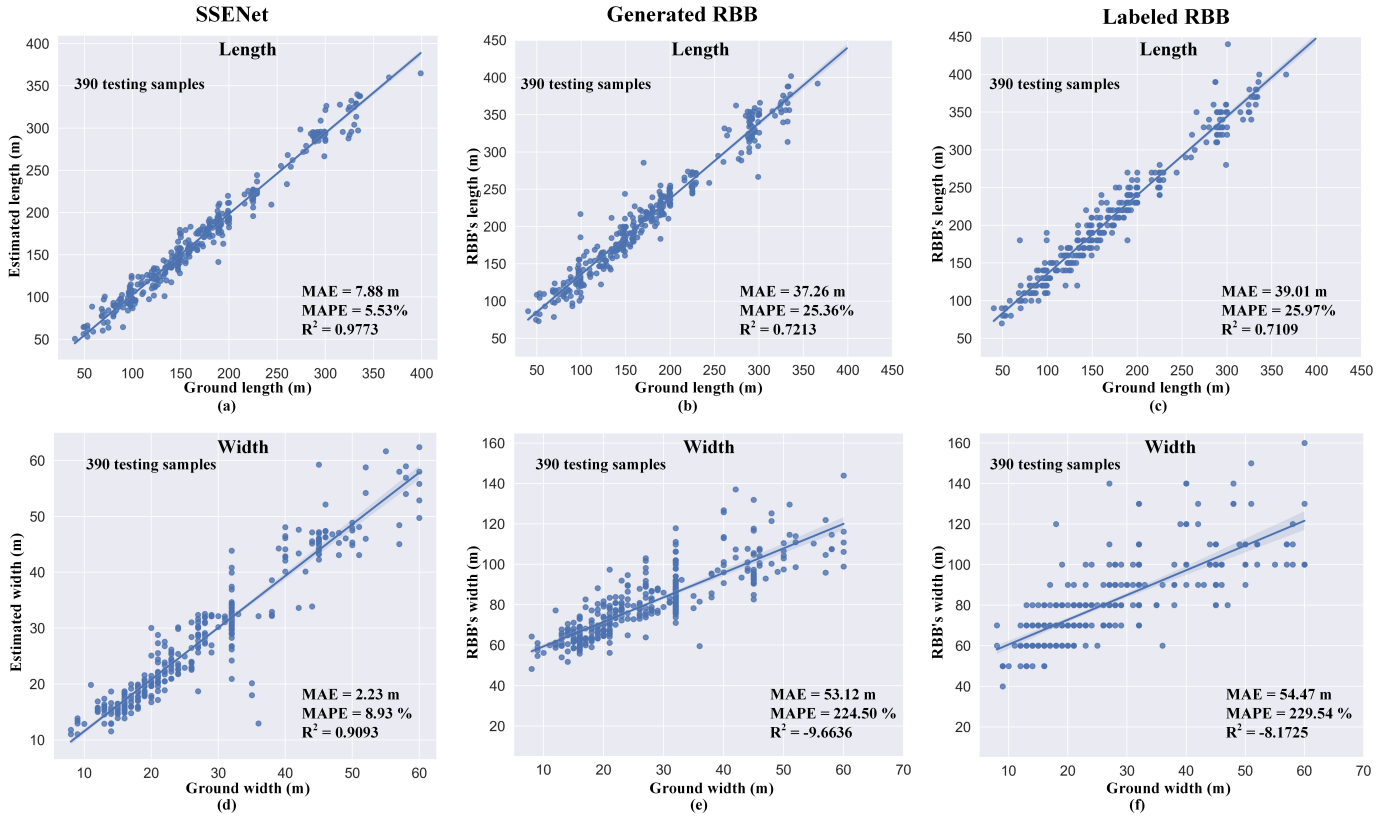


Fig. 6. Relationships between the ground size and the estimated size of SSENNet, the generated RBB's size, and the labeled RBB's size. (a)–(c) Relationships between the ground length and SSENNet's length, the generated RBB's length, and the labeled RBB's length. (d)–(f) Relationships between the ground width and SSENNet's width, the generated RBB's width, and the labeled RBB's width.

TABLE I
MODEL PERFORMANCE WITH DIFFERENT INPUTS

Models	Inputs of DNN	MAE (m)		MAPE (%)	
		Length	Width	Length	Width
SSENNet ₁	X_l, X_w	10.01	2.65	6.87	9.89
SSENNet ₂	X_l, X_w, F_6	8.14	2.27	5.82	9.22
SSENNet ₃	$X_l, X_w, \text{Cos } \theta, F_6$	7.88	2.23	5.53	8.93

that of width is more than 50 m. The gap between the RBB's size (whether generated or manually labeled) and the ship's ground size is huge. By adding the DNN-based regression model, the MAEs are pushed below 8 m, which demonstrated that the DNN-based model as the second stage of SSENNet is necessary and effective.

In summary, SSENNet can extract the ship size from SAR images end to end and control the absolute error (AE) under 0.8-pixel spacing.

D. Effectiveness of the Hybrid Input

To test the effectiveness of the hybrid input for the DNN-based regression model, we design an experiment to test the performance of SSENNet in different inputs (Table I). SSENNet₁ uses the initial ship length and width as the input without any feature map F_6 . SSENNet₂ uses the initial ship

length, ship width, and feature map as the input. SSENNet₃ uses the initial ship length, ship width, feature map, and initial orientation as the hybrid input. The other settings are unchanged.

The results are shown in Table I and Fig. 7. The MAE and MAPE of SSENNet₁ are the largest ones among the three models. By adding a feature map, SSENNet₂ obtains much better results than SSENNet₁: reducing the MAE of length by nearly 2 m. As shown in Fig. 7(a) and (b), the R^2 of SSENNet₂ is larger than that of SSENNet₁. The estimated lengths of SSENNet₂ are more consistent with the true values than those of SSENNet₁. This fact demonstrates that the input SAR image's feature map is an essential factor for ship size estimation. As stated in Section III-B, the input SAR image contains the ship's signature and the sea clutter under typical environmental conditions. The ship's signature reflects the ship's state, moving or stationary. The environmental conditions such as wind fronts, ocean waves, and rain cells affect the sea-ship interaction. All these factors are related to the ship's size and are encoded into the extracted feature map. Therefore, using the feature map as an input improves the accuracy of size estimation. Finally, by explicitly adding the ship's initial orientation as another input, the model performance is further enhanced: the MAE of length is under 8 m, which is less than 0.8-pixel spacing. Therefore, the proposed hybrid input effectively contributes to an accurate estimation.

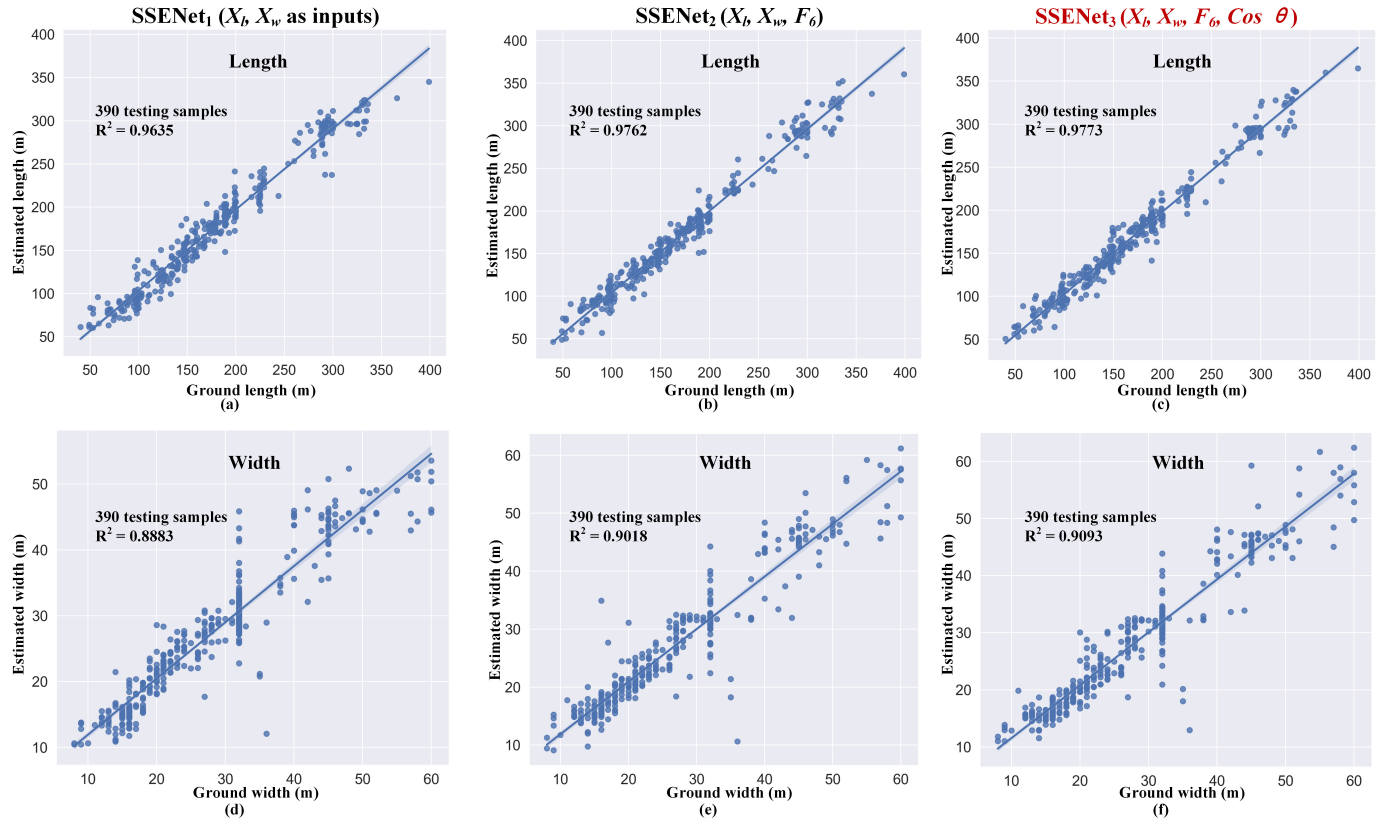


Fig. 7. Model performances with different inputs. (a)–(c) Estimated lengths and their corresponding ground truths of SSENet₁, SSENet₂, and SSENet₃. (d)–(f) Estimated widths and their corresponding ground truths of SSENet₁, SSENet₂, and SSENet₃.

TABLE II
MODEL PERFORMANCE WITH DIFFERENT LOSS FUNCTIONS

Models	Loss Function	MAE (m)		MAPE (%)	
		Length	Width	Length	Width
SSENet _{MSE}	MSE	8.85	2.20	5.99	8.81
SSENet _{MSSE}	MSSE	7.88	2.23	5.53	8.93

E. Effectiveness of MSSE Loss Function

We integrate the ground truths of the ship length and width into the classical MSE to form the MSSE loss function. To test the effectiveness of MSSE, we compare our model with the model using MSE as the loss function. The model with MSE as the loss function is named SSENet_{MSE}. SSENet_{MSSE} is the proposed model that uses MSSE as the loss function. The structure of the two models is the same except for the loss function. The other experimental settings are unchanged.

The results are shown in Table II. For length, the MAE and MAPE of SSENet_{MSSE} show apparent advantages over those of SSENet_{MSE}. The length MAE of SSENet_{MSSE} is nearly 1 m less than that of SSENet_{MSE}, an increase by 11%. Since MSSE magnifies the loss of considerable value and drives the model to focus more on length than on width during training, SSENet_{MSSE} performs slightly worse than SSENet_{MSE}. However, this drawback does not obscure the advantages of MSSE. First, the difference between the two widths is a negligible few centimeters. Second, length is a

more direct parameter to reflect the ship information compared with width. The MSSE loss improves length estimation. Thus, our MSSE loss is useful, especially for estimating the ship's length.

F. DL Model Tuning Hyperparameters

Since tuning hyperparameters is an essential procedure for the DL model, we detail this process of four key parameters, including the neuron and layer numbers of hidden NN layers, the channel numbers (N), and the max-pooling size (S) of F_6 in Section III-B.

1) *Number of Neurons in the Hidden Layers*: For the DNN regression part, the number of neurons in each hidden NN layer influences the model performance. We keep the depth of DNN as five (three hidden NN layers, one input layer, and one output layer) and change the neuron numbers of hidden NN layers to 64, 128, 256, and 512. The other settings are unchanged. The experimental result is shown in Fig. 8(a). For both length and width, MAEs and MAPEs decrease first and then increase as the number of neurons increases and reach the minimum when the number of neurons is 256. This result is consistent with typical NN models' characteristics: a small number of neurons do not fit the data adequately, and too many neurons lead to overfitting. The width of the MAPE fluctuates slightly and is minimal at 128 neurons. For most ship sizes, width is much smaller than length, which will magnify the MAPE and easily cause some fluctuations.

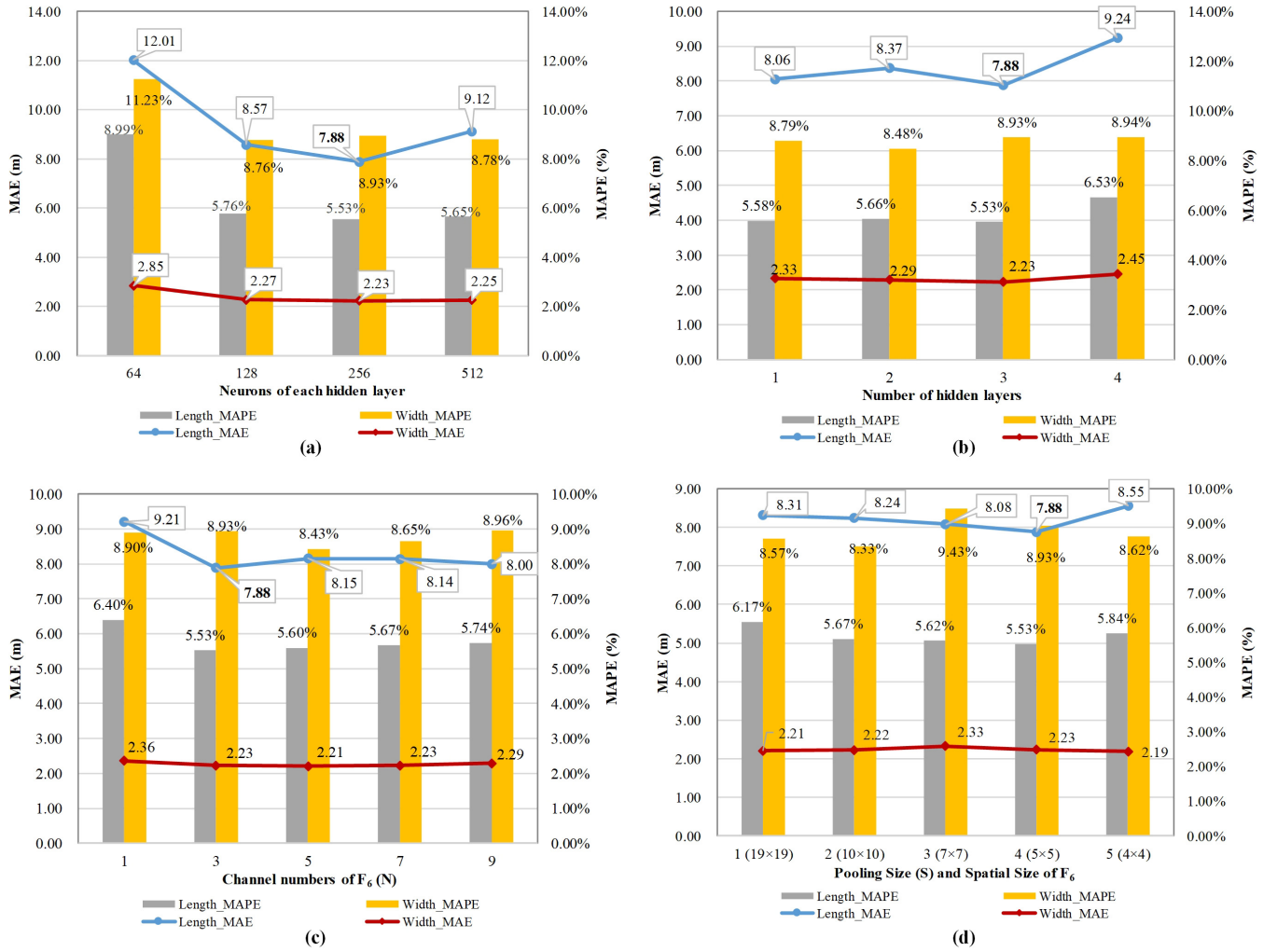


Fig. 8. Model performances of tuning hyperparameters. (a) Number of neurons in hidden layers. (b) Number of hidden layers. (c) Channel numbers of F_6 (N). (d) Pooling size of F_6 (S).

Therefore, the appropriate number of neurons for the DNN model is 256.

2) *Number of Hidden Layers*: This section tries to find the best setting for the number of hidden layers in the DNN regression part. We fix the number of neurons in all hidden layers as 256 and set the number of hidden layers as one to four in step one. The MAEs and MAPEs are shown in Fig. 8(b). When there are three hidden layers, the MAE of length, MAPE of length, and MAE of width obtain the minimum values: 7.88 m, 5.53%, and 2.23 m. When the number of layers is fewer than 3, the model performs slightly worse. This phenomenon may be due to the insufficient fitting ability of the shallow NN models. Fewer hidden layers mean fewer parameters and fewer nonlinear transformations (activation function).

For example, the one-hidden-layer model parameters are about 130000 ($256 \times 256 \times 2$) less than those of the three-hidden-layer model. The nonlinear transformations of the one-hidden-layer model are two times less than those of the three-hidden-layer model. Therefore, the fitting ability of one-hidden-layer-model is weaker than that of the three-hidden-layer mode, and the same is true for the two-hidden-layer-model. When the number of layers is 4, the performance

of the model is low. This low performance may due to the gradient vanishing in a deeper model. The core idea of DL is to increase the nonlinear transformation capability of a model by employing hidden NN layers as many as possible, the so-called a more in-depth model. However, too many hidden NN layers lead to a gradient vanishing problem, making the model difficult to be trained and reducing the model's performance. Therefore, the hidden layers of a DNN model cannot be increased indefinitely.

3) *Channel Numbers of F_6 (N)*: In Section III-B, we compress F_5 in channel dimension by a CNN layer with $N 1 \times 1$ filters. This experiment explores the effects of different N values on the performance of the size estimation. We set the max-pooling layer's kernel size, defined as S in Section III-B, to 4. Then, the spatial dimension of F_6 is 5×5 . The other settings are unchanged. The results are shown in Fig. 8(c). For length estimation, the model does not perform very well when N is 1 because the MAE is 9.21 m. When N has other values, the estimation errors do not differ much, reaching a minimum of 7.88 m when N is 3. For width estimation, different N values show little impact on MAEs. Therefore, we set the value of N to 3 in our proposed model.

TABLE III
APPROPRIATE VALUES OF FOUR HYPERPARAMETERS

Number of Neurons	Number of Hidden Layers	The Channel Numbers of F_6 (N)	The Pooling Size of F_6 (S)
256	3	3	4

4) *Pooling Size of F_6 (S)*: In Section III-B, we compress F_5 in spatial dimension by a max-pooling layer with $S \times S$ filters. This experiment explores the effects of different S values on the performance of the size estimation. We fix the channel numbers (N) to 3 and change S from 1 to 5 in step 1. As the spatial size of F_5 is 19×19 , the spatial dimension of F_6 can be calculated by $\lceil 19/S \rceil$. Therefore, the spatial dimension of F_6 corresponding to filters with sizes 1, 2, 3, 4, and 5 is 19×19 , 10×10 , 7×7 , 5×5 , and 4×4 . The results are shown in Fig. 8(d). For length estimation, the MAE of length fluctuates slightly with different values of S and is minimum when S is 4, which is 7.88 m. The MAE of width changes very little. Therefore, we set the value of S as 4 in our proposed model.

Finally, the appropriate values of four hyperparameters are shown in Table III.

V. DISCUSSIONS

A. ML Versus DL in Practical Ship Size Estimation?

The accurate ship size regression part in the proposed model is a DNN model. We chose three typical ML models—gradient boosting regression (GBR) [46], support vector regression (SVR) [47], and linear regression (LR) [48]—and discuss their performances on our experimental data. GBR and SVR are two typical nonlinear models that are adopted by existing ship size extraction models in references [8] and [7]. LR is a basic linear model used in reference [13]. Because these three ML models cannot be integrated with the SSD to form an end-to-end model, we manually labeled the RBB of each ship as the initial ship length, width, and ship orientation and inputted them into three ML models. The training set and testing are unchanged. We tune the parameters of GBR, SVR, and LR and record their results with minimum MAEs. The DNN model is the one integrated into the proposed DL model.

The results are shown in Table IV, with our DL model results in the last row for reference. For length estimation, GBR obtains the minimum MAE, that is, 9.69 m. For width estimation, GBR also performs the best. LR, SVR, and DNN perform worse than GBR.

The GBR works the best. GBR is a typical ensemble learning model that achieves better results than a single learner by training and combining multiple learners. GBR has been proved to be the right choice in the three-stage ship size estimation procedure [8]. However, since GBR is an ML model based on decision trees, it cannot automatically extract features from input SAR images and integrally perform the traditional three steps. It is also not possible to integrate GBR with a DL-based ship detection model, such as DRBox-v2. The premise of using GBR is that we need to accurately extract

TABLE IV
MAE AND MAPE \pm CONFIDENCE INTERVAL (CI) OF DIFFERENT MODELS

Models	Inputs	MAE (m)		MAPE (%)	
		Length	Width	Length	Width
LR	RBB	$10.60 \pm 1.05^*$	2.86 ± 0.29	6.97 ± 0.86	10.47 ± 0.91
SVR	RBB	10.16 ± 1.06	2.48 ± 0.32	6.61 ± 0.84	9.50 ± 1.14
GBR	RBB	9.69 ± 0.96	2.39 ± 0.29	6.93 ± 0.91	9.27 ± 1.10
DNN	RBB	10.79 ± 1.04	2.83 ± 0.29	7.27 ± 0.92	10.22 ± 0.98
SSENet	Images	7.88 ± 0.77	2.23 ± 0.27	5.53 ± 0.65	8.93 ± 1.04

* $CI = 1.96 \times \sigma(x) / \sqrt{N}$, where σ is the standard deviation of x (for MAE, x is absolute error of each sample, for MAPE x is percentage absolute error of each sample) and N is the number of ships in the testing set (390). All values are statistically significant at the 95% confidence level.

the RBB of the ship by binarization and image operation, which is a great challenge, especially in the era of big data. Practically, GBR is unable to achieve an end-to-end size extraction: inputting the SAR image and outputting the ship size.

Table IV shows that the DNN model does not perform well. However, the advantage is that we can integrate a DNN model with any DL-based ship detection model based on CNN or NN to achieve an end-to-end size extraction from the SAR image. Compared with the traditional three-stage method, all trainable parameters in the DL model are globally optimized. The feature maps automatically extracted by the DL detection model can be input into the DNN regression model to further improve ship size's extraction accuracy. As shown in Table IV, compared with the GBR model, our SSENet reduces the MAE of length by nearly 2 m, that is, by about 18.68%. Therefore, we design a DNN model for accurate ship size regression.

B. Sources of the Errors?

This section analyzes the estimation errors in detail and tries to find out what causes large errors. Based on existing studies [8], [14], we mainly analyze the relationship between MAE error and ship orientation and travel speed.

1) *Ship Orientation*: In Fig. 9, we display the trend of errors with respect to the orientation angle. The orientation angles are ground values obtained from AIS. Fig. 9(a) and (c) is the results of the model without orientation as an input parameter (SSENet₂ in Table I). Fig. 9(b) and (d) is the results of the model with orientation as an input parameter (SSENet₃ in Table I). For both models, the error varies with the ship orientation variation. The errors are higher when the orientation angles are closer to 0° (azimuth direction) in the range of $(-45^\circ, 45^\circ)$. This phenomenon is mainly caused by the ship motion and the unequal resolution after image formation ($5 \text{ m} \times 20 \text{ m}$ for range and azimuth directions, respectively; the lower resolution enlarges the errors in the azimuth direction) [8]. If the ship moves in a direction consistent with the azimuthal direction, the azimuth direction's speed component is large. The large speed component leads to greater smearing of ship signature in the SAR image, which increases the estimation error.

By comparing Fig. 9(a) and (b) with Fig. 9(c) and (d), we find that when we add the initial orientation angle ($\cos \theta$)

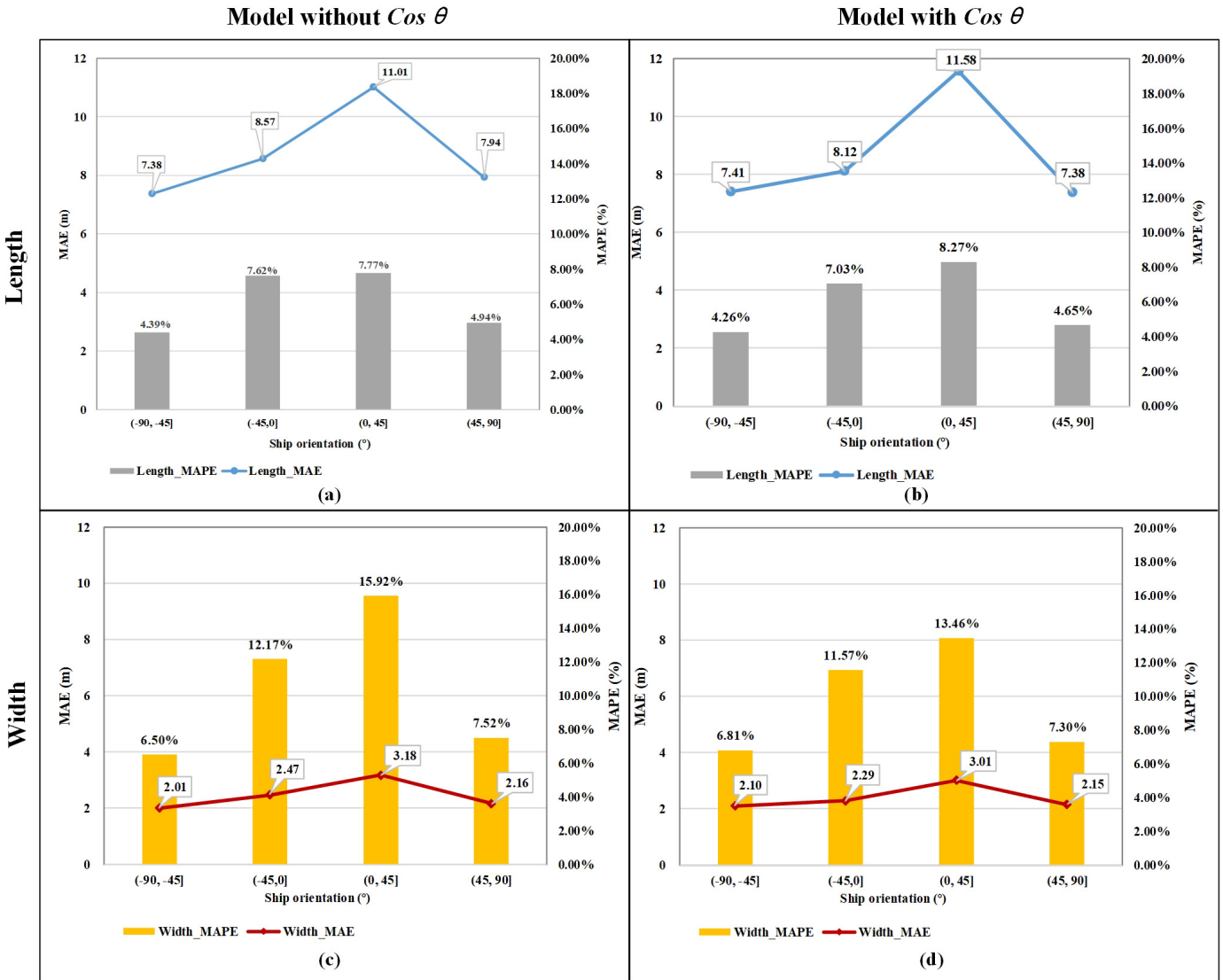


Fig. 9. Trend of errors with respect to the orientation angle. (a) Length error of the model without $\cos \theta$. (b) Length error of the model width $\cos \theta$ system. (c) Width error of the model without $\cos \theta$. (d) Width error of the model without $\cos \theta$.

into the DNN model as an input, although the overall trend does not change, the error in most angle intervals is reduced. This result further verifies the validity of adding the initial orientation angle as an input in the DNN regression model.

2) *Ship Speed*: Fig. 10 shows the trend of errors in the ship's speed. As the SAR images of the OpenSARShip are mainly from ports, about 83% of the ships are still. Fig. 10(a) and (b) shows that the MAEs of length and width in the range of (0, 1) knot (1852 km/h) are small. The error fluctuates as the ship speed increases. When the speed is greater than 15 kn (27780 km/h), the MAEs of length and width increase significantly, which are 19.04 and 4.71 m, respectively, far greater than the errors of other speed intervals. Thus, when the ship's speed is greater than 15 kn (27780 km/h), the AE in size estimation significantly increases. However, unlike the ship's initial length, width, and orientation, we cannot obtain the reliable ship speed from its signature on the SAR image. One may get ship speed from its wake features in the SAR image [49]. However, systematically obtaining ship speed from

SAR images under different environmental conditions still needs further research.

3) *Ship Size*: Fig. 11 shows the absolute and relative errors of each predicted size to the ground ship size. The AE of a predicted size is calculated by setting N in MAE as 1. The relative error is the percentage error (PE), which is obtained by setting N in MAPE as 1. As shown in Fig. 11(a) and (c), there are no apparent dependencies between the AE of the predicted size and the ground size. The PE goes down with the increase in the ground size see [Fig. 11(b) and (d)]. This phenomenon is easy to understand because an increase in the ground value leads to an increase in the PE denominator, resulting in a decrease in the PE. Therefore, the ship size is not a source of errors.

C. Robustness

To explore the robustness of SSENNet, we test SSENNet by the leave-one-out strategy. There are 1890 samples. We construct

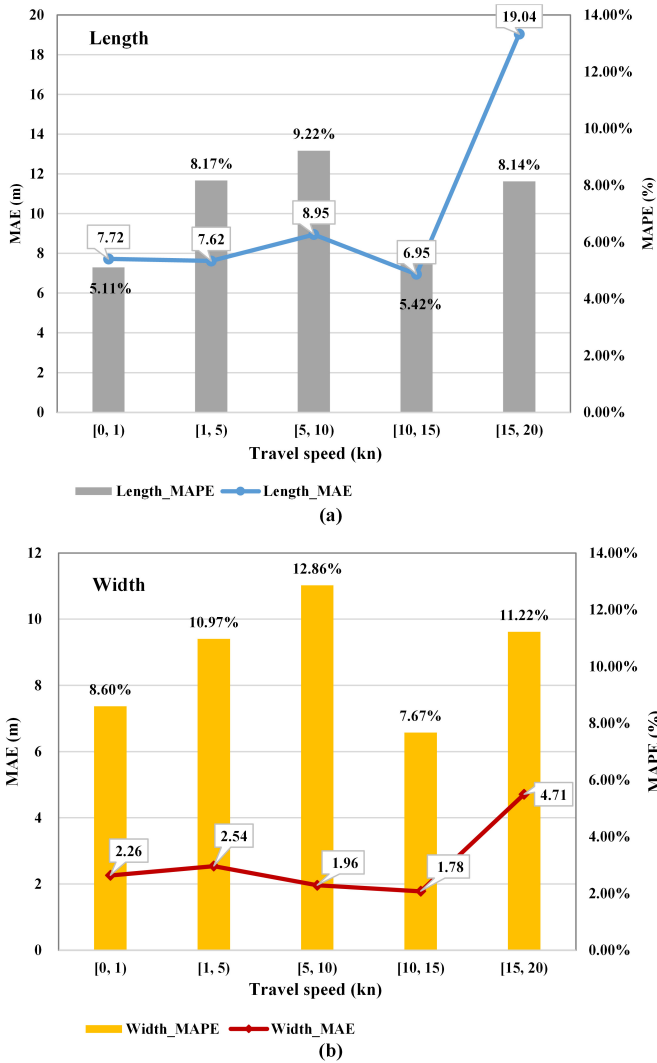


Fig. 10. Trend of errors with respect to the travel velocity. (a) Trend of length error. (b) Trend of width error.

TABLE V
PERFORMANCES ON DIFFERENT TESTING SETS

Data	MAE (m)		MAPE (%)	
	Length	Width	Length	Width
Data ₁	7.88	2.23	5.53	8.93
Data ₂	8.34	2.29	6.43	10.66
Data ₃	7.73	2.02	5.57	8.49
Data ₄	7.90	1.87	5.90	8.46
Mean	7.96	2.10	5.86	9.14

four testing sets named Data₁, Data₂, Data₃, and Data₄. Each testing set contains 390 samples, and the samples between different test sets do not overlap. Each time, we selected one of the four testing sets as the testing samples and the remaining 1500 samples as the training set. Data₁ is the testing set we used to evaluate the model performance and discuss the model's errors in Sections IV and V-A and V-B. The testing results are shown in Table V. For length, the MAEs of the three testing sets (Data₁, Data₃, and Data₄) are less than 8 m

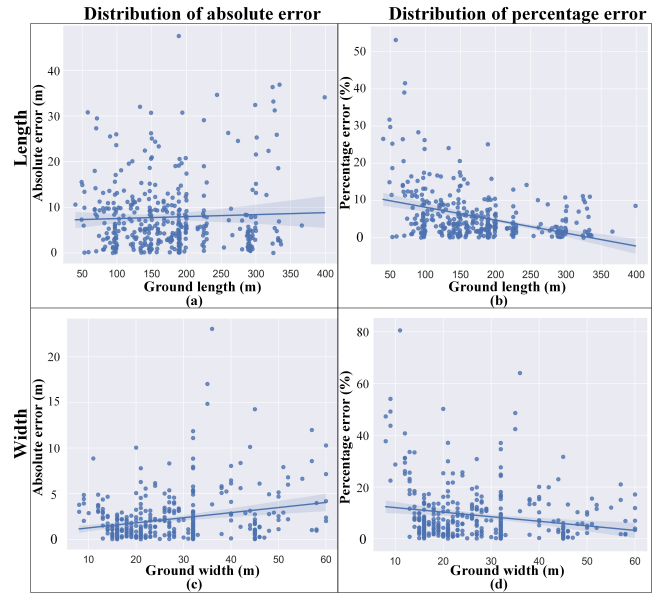


Fig. 11. Distribution of absolute and relative errors with respect to the ship size. (a) Absolute errors of length (b) Relative errors (percentage error) of length. (c) Absolute errors of width. (d) Relative errors (percentage error) of width.

(0.8-pixel), and the mean MAE of four testing sets is 7.96 m. For width, the MAEs of all testing sets are under 2.30 m, and the mean MAE is 2.10 m. Therefore, SSENNet shows robustness on different training/testing sets.

VI. CONCLUSION

This study proposes an end-to-end DL model, SSENNet, to extract the ship size from SAR images. We employ an SSD-based model, DRBox-v2, to extract high-level features and generate an RBB of the ship from the input SAR image. Based on the extracted features, we design a DNN-based regression model to estimate the accurate ship size. The DNN-based regression model is concatenated with the SSD-based model to form SSENNet. We construct a hybrid input for the DNN-based regression model. The hybrid input includes the initial ship size and orientation angle obtained from the RBB and the high-level features extracted from the input SAR image. To characterize length and width of a ship accurately, we design a custom loss function MSSE to optimize the DNN model. SSENNet is trained and validated by the OpenSARShip, a data set dedicated to Sentinel-1 ship interpretation. Experiments show that: 1) SSENNet is capable of extracting ship size from SAR images end to end and pushes the estimation error MAE under 0.8-pixel, 7.88 m in length and 2.23 m in width; 2) compared with the single input, the hybrid input significantly reduces the estimation errors, about 2 m in length and 0.4 m in width; 3) the MSSE reduces the MAE of length by nearly 1 m and increases the MAE of width 0.03 m compared to the classical MSE loss function; 4) compared with the typical traditional ML model GBR, SSENNet simplifies the extraction procedure and reduces the MAE of length by nearly 2 m (18.68%) and the MAE of width by 0.06 m (2.51%), showing advantages of the end-to-end learning; and 5) SSENNet

shows robustness on four different training/testing sets, with a mean MAE of length 7.96 m and a mean MAE of width 2.10 m.

As the ship orientation and travel speed are two main factors impacting the accuracy of size estimation, in the future, we will explore how to optimize the model to eliminate the influence of these two factors. Based on the extracted size, we will also attempt to develop a DL model to identify ship types end to end from SAR images. To test the model's adaptability to different SAR sensors, we will try to build more SAR ship data sets and make some quantitative evaluations.

ACKNOWLEDGMENT

The OpenSARShip is downloaded from "OpenSAR Platform" (<http://opensar.sjtu.edu.cn>).

REFERENCES

- [1] M. F. Fingas and C. E. Brown, "Review of ship detection from airborne platforms," *Can. J. Remote Sens.*, vol. 27, no. 4, pp. 379–385, Aug. 2001.
- [2] R. Torres *et al.*, "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.
- [3] T. Zhang, J. Ji, X. Li, W. Yu, and H. Xiong, "Ship detection from PolSAR imagery using the complete polarimetric covariance difference matrix," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2824–2839, May 2019, doi: [10.1109/tgrs.2018.2877821](https://doi.org/10.1109/tgrs.2018.2877821).
- [4] K. Eldhuset, "An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1010–1019, Jul. 1996.
- [5] K. Ouchi, S. Tamaki, H. Yaguchi, and M. Iehara, "Ship detection based on coherence images derived from cross correlation of multilook SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 3, pp. 184–187, Jul. 2004.
- [6] H. Lang, Y. Xi, and X. Zhang, "Ship detection in high-resolution SAR images by clustering spatially enhanced pixel descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5407–5423, Aug. 2019, doi: [10.1109/tgrs.2019.2899337](https://doi.org/10.1109/tgrs.2019.2899337).
- [7] H. Lang and S. Wu, "Ship classification in moderate-resolution SAR image by naive geometric features-combined multiple kernel learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1765–1769, Oct. 2017, doi: [10.1109/lgrs.2017.2734889](https://doi.org/10.1109/lgrs.2017.2734889).
- [8] B. Li, B. Liu, W. Guo, Z. Zhang, and W. Yu, "Ship size extraction for Sentinel-1 images based on dual-polarization fusion and nonlinear regression: Push error under one pixel," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4887–4905, Aug. 2018, doi: [10.1109/tgrs.2018.2841882](https://doi.org/10.1109/tgrs.2018.2841882).
- [9] M. Stasolla and H. Greidanus, "The exploitation of Sentinel-1 images for vessel size estimation," *Remote Sens. Lett.*, vol. 7, no. 12, pp. 1219–1228, Dec. 2016, doi: [10.1080/2150704x.2016.1226522](https://doi.org/10.1080/2150704x.2016.1226522).
- [10] C. C. Wackerman, K. S. Friedman, W. G. Pichel, P. Clemente-Colón, and X. Li, "Automatic detection of ships in RADARSAT-1 SAR imagery," *Can. J. Remote Sens.*, vol. 27, no. 4, pp. 371–378, Aug. 2001.
- [11] C. Wang, M. Liao, and X. Li, "Ship detection in SAR image based on the alpha-stable distribution," *Sensors*, vol. 8, no. 8, pp. 4948–4960, Aug. 2008.
- [12] X. Qin, S. Zhou, H. Zou, and G. Gao, "A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 806–810, Jul. 2013.
- [13] B. Tings, C. A. Bentes da Silva, and S. Lehner, "Dynamically adapted ship parameter estimation using TerraSAR-X images," *Int. J. Remote Sens.*, vol. 37, no. 9, pp. 1990–2015, May 2016, doi: [10.1080/01431161.2015.1071898](https://doi.org/10.1080/01431161.2015.1071898).
- [14] L. Huang *et al.*, "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018, doi: [10.1109/jstars.2017.2755672](https://doi.org/10.1109/jstars.2017.2755672).
- [15] S. Ben Taieb and R. J. Hyndman, "A gradient boosting approach to the kaggle load forecasting competition," *Int. J. Forecasting*, vol. 30, no. 2, pp. 382–394, Apr. 2014.
- [16] R. Pelich, N. Longepe, G. Mercier, G. Hajdуч, and R. Garello, "Performance evaluation of Sentinel-1 data in SAR ship detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Milan, Italy, Jul. 2015, pp. 2103–2106, doi: [10.1109/IGARSS.2015.7326217](https://doi.org/10.1109/IGARSS.2015.7326217).
- [17] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
- [19] B. Liu, X. Li, and G. Zheng, "Coastal inundation mapping from bitemporal and dual-polarization SAR imagery based on deep convolutional neural networks," *J. Geophys. Res.: Oceans*, vol. 124, no. 12, pp. 9101–9113, Dec. 2019, doi: [10.1029/2019JC015577](https://doi.org/10.1029/2019JC015577).
- [20] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.
- [21] Z. I. Petrov and Y. Tian, "Prediction of sea ice motion with convolutional long short-term memory networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6865–6876, Sep. 2019, doi: [10.1109/tgrs.2019.2909057](https://doi.org/10.1109/tgrs.2019.2909057).
- [22] X. Li *et al.*, "Deep-learning-based information mining from ocean remote-sensing imagery," *Nat. Sci. Rev.*, vol. 7, no. 10, pp. 1585–1606, 2020, doi: [10.1093/nsr/nwaa047](https://doi.org/10.1093/nsr/nwaa047).
- [23] Y. Ren, T. Cheng, and Y. Zhang, "Deep spatio-temporal residual neural networks for road-network-based data modeling," *Int. J. Geographical Inf. Sci.*, vol. 33, no. 9, pp. 1894–1912, Sep. 2019, doi: [10.1080/13658816.2019.1599895](https://doi.org/10.1080/13658816.2019.1599895).
- [24] G. Zheng, X. Li, R. H. Zhang, and B. Liu, "Purely satellite data-driven deep learning forecast of complicated tropical instability waves," *Sci. Adv.*, vol. 6, no. 29, 2020, Art. no. eaba1482, doi: [10.1126/sciadv.aba1482](https://doi.org/10.1126/sciadv.aba1482).
- [25] X. Zhang and X. Li, "Combination of satellite observations and machine learning method for internal wave forecast in the sulu and cebeles seas," *IEEE Trans. Geosci. Remote Sens.*, early access, Jul. 16, 2020, doi: [10.1109/TGRS.2020.3008067](https://doi.org/10.1109/TGRS.2020.3008067).
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
- [27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [28] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019, doi: [10.3390/rs11070765](https://doi.org/10.3390/rs11070765).
- [29] 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, doi: [10.1109/LGRS.2018.2882551](https://doi.org/10.1109/LGRS.2018.2882551).
- [30] Y. Wang, C. Wang, and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images," *Remote Sens. Lett.*, vol. 9, no. 8, pp. 780–788, Aug. 2018, doi: [10.1080/2150704x.2018.1475770](https://doi.org/10.1080/2150704x.2018.1475770).
- [31] X. Zhang *et al.*, "A lightweight feature optimizing network for ship detection in SAR image," *IEEE Access*, vol. 7, pp. 141662–141678, 2019, doi: [10.1109/ACCESS.2019.2943241](https://doi.org/10.1109/ACCESS.2019.2943241).
- [32] Y.-L. Chang, A. Anagaw, L. Chang, Y. Wang, C.-Y. Hsiao, and W.-H. Lee, "Ship detection based on YOLOv2 for SAR imagery," *Remote Sens.*, vol. 11, no. 7, p. 786, Apr. 2019, doi: [10.3390/rs11070786](https://doi.org/10.3390/rs11070786).
- [33] Z. Deng, H. Sun, S. Zhou, and J. Zhao, "Learning deep ship detector in SAR images from scratch," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019, doi: [10.1109/tgrs.2018.2889353](https://doi.org/10.1109/tgrs.2018.2889353).
- [34] T. Zhang, X. Zhang, J. Shi, and S. Wei, "HyperLi-net: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 123–153, Sep. 2020, doi: [10.1016/j.isprsjrs.2020.05.016](https://doi.org/10.1016/j.isprsjrs.2020.05.016).
- [35] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2016, pp. 21–37.
- [36] L. Liu, Z. Pan, and B. Lei, "Learning a rotation invariant detector with rotatable bounding box," 2017, *arXiv:1711.09405*. [Online]. Available: <https://arxiv.org/abs/1711.09405>
- [37] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019, doi: [10.1109/tgrs.2019.2920534](https://doi.org/10.1109/tgrs.2019.2920534).
- [38] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

- [39] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [40] X. Yang *et al.*, "Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, p. 132, Jan. 2018.
- [41] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: <http://arxiv.org/abs/1409.0473>
- [42] G. Margarit, J. J. Mallorqui, J. Fortuny-Guasch, and C. Lopez-Martinez, "Study of the influence of vessel motions and sea-ship interaction on classification algorithms based on single-pass polarimetric SAR interferometry," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, Jul. 2006, pp. 75–78.
- [43] G. Margarit, J. J. Mallorqui, J. Fortuny-Guasch, and C. Lopez-Martinez, "Exploitation of ship scattering in polarimetric SAR for an improved classification under high clutter conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1224–1235, Apr. 2009.
- [44] P. Liu and Y.-Q. Jin, "A study of ship rotation effects on SAR image," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3132–3144, Jun. 2017.
- [45] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 91–99.
- [46] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232 and 44, 2001, doi: 10.1214/aos/1013203451.
- [47] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 9, 1996, pp. 155–161.
- [48] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: Wiley, 2012.
- [49] C. R. Jackson and J. R. Apel, *Synthetic Aperture Radar Marine User's Manual*. Washington, DC, USA: National Oceanic and Atmospheric Administration, 2004.



Yibin Ren (Member, IEEE) received the B.S. degree in geographical information science from the Shandong University of Science and Technology, Qingdao, China, in 2012, the M.S. degree in cartography and geography information system from the Nanjing University, Nanjing, China, in 2015, and the joint Ph.D. degree in cartography and geography information system from the Ocean University of China, Qingdao, and University College London, London, U.K., in 2019.

He has been holding a post-doctoral position with the Institute of Oceanology, Chinese Academy of Sciences, Qingdao, since 2019. His research interests include target detection and semantic segmentation from ocean remote sensing data, and ocean prediction based on deep learning.



Xiaofeng Li (Fellow, IEEE) received the B.S. degree in optical engineering from Zhejiang University, Hangzhou, China, in 1985, the M.S. degree in physical oceanography from the First Institute of Oceanography, Qingdao, China, in 1992, and the Ph.D. degree in physical oceanography from North Carolina State University, Raleigh, NC, USA, in 1997.

From 1997 to 2019, he was with the National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, USA. His research interests include synthetic aperture radar (SAR) applications in oceanography and marine meteorology, artificial intelligence oceanography, big data, and satellite image processing.

Dr. Li is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the *International Journal of Remote Sensing*. He is an Editorial Board Member of the *International Journal of Digital Earth*, *Big Earth Data*, and the *Journal of Oceanology and Limnology*. He is also the Executive Editor-in-Chief of the *Journal of Remote Sensing* (a Science Partner Journal).



Huan Xu (Student Member, IEEE) received the B.S. degree in surveying and mapping engineering from the Jiangsu Ocean University, Lianyungang, China, in 2018. He is pursuing the joint M.S. degree in surveying and mapping engineering with Jiangsu Ocean University and the Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China.

His research interests include target detection and classification from ocean SAR data based on deep learning.