Enhancement of Ship Type Classification from a Combination of CNN and KNN

Ho-Kun Jeon 1,2 and Chan-Su Yang 1,2,3,*

1 Marine Security and Safety Research Center, Korea Institute of Ocean Science & Technology, Busan 49111, Korea; hkjeon@kiost.ac.kr
2 Department of Ocean Science, University of Science & Technology, Daejeon 34113, Korea
3 Department of Convergence Study on the Ocean Science and Technology, Ocean Science and Technology School, Korea Maritime and Ocean University, Busan 49111, Korea
* Correspondence: yangcs@kiost.ac.kr; Tel.: +82-10-3478-8637

Abstract: Ship type classification of synthetic aperture radar imagery with convolution neural network (CNN) has been faced with insufficient labeled datasets, unoptimized and noised polarization images that can deteriorate a classification performance. Meanwhile, numerous labeled text information for ships, such as length and breadth, can be easily obtained from various sources and can be utilized in a classification with k-nearest neighbor (KNN). This study proposes a method to improve the efficiency of ship type classification from Sentinel-1 dual-polarization data with 10 m pixel spacing using both CNN and KNN models. In the first stage, Sentinel-1 intensity images centered on ship positions were used in a rectangular shape to apply an image processing procedure such as head-up, padding and image augmentation. The process increased the accuracy by 33.0% and 31.7% for VH (vertical transmit and horizontal receive) and VV (vertical transmit and vertical receive) polarization compared to the CNN-based classification with original ship images, respectively. In the second step, a combined method of CNN and KNN was compared with a CNN-alone case. The f1-score of CNN alone was up to 85.0%, whereas the combination method showed up to 94.3%, which was a 9.3% increase. In the future, more details on an optimization method will be investigated through field experiments of ship classification.

Keywords: ship classification; CNN; KNN; Sentinel-1

1. Introduction

Marine surveillance has become a crucial issue due to illegal activity in the oceans [1,2]. To counteract the threat, space-borne remote sensing technology has been used for its powerful observability over a wide area. Among them optical remote sensing technology had shown its availability to detect objects for their virtue of high reflectance characteristics though, their detection performance is constrained by meteorological phenomena and atmospheric conditions such as cloud and night [3]. Meanwhile, space-borne synthetic aperture radar (SAR) imagery has become an effective material to observe targets regardless of weather conditions above [4,5]. The increasing number of launching SAR satellites has increased the accessibility of monitoring objects on the sea without high costed radar and patrol fleet through open-public data such as RADARSAT and Sentinel-1 of the European space agency (ESA). Accordingly, research on object detection and classification using SAR imagery has been developed with image preprocessing and deep learning methods.

Image preprocessing is the significant stage to design data standardized and represents targets’ features so that the images are used as training datasets. Quadrature and equal-sized images are a standardized form to input recognizable data into a convolution network model [6,7]. Target centroid centralization enhances the comparison between different images, making the model recognize the target easily [8]. Data argumentation is also required to solve the lack of training datasets; to address this, rotation [8] and
multiview combination [9] have been devised. Generation of dual-polarization images such as reflection symmetry metric (RSM) and PMA is also helpful to make the vivid contrast between targets and surrounding noises [10,11].

Among deep learning methods, a convolution neural network (CNN) has been widely used in image classification [12] and developed into various models [13–15]. CNN applied various fields such as sea fog recognition [16], traffic signal recognition [17] and face recognition [18].

Precedent research on CNN trained with SAR imagery has introduced the method to improve the classification performance. Gao et al. (2016) computed intra-class and inter-class distance and utilized them for cost reduction [19]. Lin et al. (2017) devised a convolution highway unit that trains a model with limited SAR data by a deeper network [20]. Lang et al. (2018) and Song et al. (2020) suggested transferring given information to a model [21,22]. Ma et al. (2018) used pyramid architecture to make the model deeper [23]. Xie et al. (2019) developed an umbrella structure to train diverse characteristics of targets from different levels [24]. Huang et al. (2018) generated ship image datasets that were calibrated radiometrically and geometrically [25].

Despite the numerous trials using SAR imagery as training datasets for CNN, there have been three common constraints in ship detection and classification. The first is the expensive cost and long visiting period. It provides a result showing that the training datasets are not enough, and provides overfitting [7,9].

The second is their distinguished swath and polarization [9]. It makes different classification results because the same ship can have various properties due to both characteristics [9,11]. The last is that discriminating shape of ships is disturbed due to the high backscattering of ocean backgrounds [26] such as wave and wake, and noises that cause false alarms [26,27]. Thus, more information about ships, comparison of classification performance with various polarization, and standardized datasets are required to compensate for the constraints.

For the first, text information that represents ship types can be considered as a supplement instead of more images. For this reason, the automatic identification system (AIS) can be considered [2,21,22,25]. AIS utilized in marine surveillance consists of dynamic and static data [28,29]. Among them, static data includes ship’s length and breadth are good features to represent ship types. For the second issue, dual-polarization images can be considered to maximize the target feature or minimize the noises, using the characteristics of VV (vertical transmit and vertical receive) and VH (vertical transmit and horizontal receive) polarization. Co-polarization VV is effective to show sea surface states due to its high backscattering value meanwhile cross-polarization VH is good for object detection because sea clutter is expressed as lower values [2,10,30].

This paper introduces the ship type prediction method from modified ship images and ship’s size information with CNN [12–24] and K-Nearest Neighbor models [31], respectively. OpenSARship, one of the SARship datasets [25], is used in this experiment. The datasets are generated from 41 Sentinel-1 (S-1) imagery that radiometrically corrected and post-processing was made by humans to obtain high-quality data, combining AIS messages.

Both CNN and KNN models compute and return each probability of three ship types such as cargo, tanker and others. Ship type is determined from four simple methods that use threshold, average and standard deviation of probability for the ship types from image and text learning.

2. Data

2.1. Modified Ship Images

SAR ship images from OpenSARship [32] were used, a dataset consisting of 11,346 single ship images and its information data on detected ships from Sentinel-1 imagery. They were radiometrically calibrated and the ship types were labeled semi-automatically. Ship information data includes ships, type, heading, length and breadth from AIS [25]. Ground range detection (GRD) images were selected because they have a 10 by 10 m spacing
distance for a pixel and it makes it simple to measure horizontal and vertical distance for modification of images.

There were three challenges to use the SAR images. Firstly, every image has a different size according to ship size. For equal training and testing, the images should be the same size. Secondly, the direction of every single ship is different. It makes a CNN model to recognize the target as a different object. Thirdly, the SAR ship images have been labeled with ship type for every file name though, the images have noise around a single ship and do not have distinctive characteristics for the ship type.

Thus, the images first were cropped into 96 by 96 pixel and rotated to make ship’s heading toward the top using ship information data, as shown in Figure 1. Then the images were categorized into three ship types as cargo, tanker, and others. Since we still found that many of them do not show type-specific characteristics, proper ship images that have discriminable shape and brightness were manually selected 100 images per ship type. Among them, 70 and 30 ship images per ship type were arranged for training and testing, respectively. The 280 training and 90 testing images were prepared, accordingly. Cargo has bright pixels on accommodation position whereas tanker has bright values on its overall hull. Others including passenger, search and rescue and tug, are relatively smaller than the cargo/tanker, as shown in Figure 1.

![Figure 1. Ship images along to type. Firstly selected 100 good quality images per ship type. Secondly, images are rotated using heading. Finally, padding is made around the ship bounding area using length and breadth.](image)

The images then were made for the ship’s heading toward the top using ship information data (head-up), as Figure 2a. But the azimuth direction of S-1 satellite is slightly tilted from the true north, the images do not exactly toward the top. Prior to full-scale classification, the effect of head-up was verified by a simple test with the 314 original SAR ship images per ship type. The simple test showed original images results in 52.6% and 51.6% for VH and VV polarization in classification accuracy. For modified images, there were 2.7% and 1.3% increases in classification accuracy for VH and VV, respectively.

![Figure 2. Modification of ship image. (a) Outer of rectangle area is padded with black color, rectangle area is computed using length and breadth information, and ship image size was fixed as 96 by 96 pixels for CNN training and testing. (b) Ship image augmentation was made using bright contrast and rotation thereafter 18 times datasets were obtained.](image)
As referred to in the introduction, the SAR ship images include noises and scattered pixels that affected by waves. Thus, the padding method is used to reduce noises around a single ship. Rectangle area is computed using length and breadth in ship information data, as shown in Figure 2a. Because the ship’s heading on the image is slightly tilted from the top, more 10 pixels were used as a buffer to make a rectangle area. To increase the number of the training datasets, the image augmentation was made by brightness contrast, rotation and flipping. The first contrast (cont1) contributes to removing noise around the ship and the second contrast (cont2) gives higher brightness to the ship’s pixels. Rotation was made with 90, 180 and 270 degrees, and additionally ship images were flipped up/down (FlipUD) and left/right (FlipLR), as shown in Figure 2b. Subsequently, 18 times the number of ship images, 1260 ($70 \times 18$) training images per ship type, are obtained. For verification of the effectiveness of the image modification above, a classification test was conducted both for the original SAR ship images and modified ship images. There were around 33.0% and 31.7% increase for VH and VV in classification accuracy from the original SAR ship images.

S-1 image has different products according to its polarizations such as VH and VV. The polarizations show different brightness for the same pixels. Aggregation of two polarization can improve classification performance [11]. Maximum and minimum images, named maxVHVV and minVHVV, are used in this study as shown in Figure 3. The maxVHVV is produced by taking a higher value for the pixel among two polarizations meanwhile minVHVV takes a lower value. Thus, the training and test images for this study are ready after three types of categorization, selection of good images, rotation, padding, and creation of maxVHVV and minVHVV.

![Figure 3](image)

**Figure 3.** Two existent polarizations and two newly made images. Newly made minVHVV takes lower values meanwhile maxVHVV takes higher values both from VV (vertical transmit and vertical receive) and VH (vertical transmit and horizontal receive) existent polarization.

### 2.2. Korean Coast Static AIS

To obtain a large number of training data, static AIS around the Coast of South Korea has been collected from 1 January 2019 to 10 February 2021, approximately two years. They are from five stations, located Busan, Socheongcho, Ulleungdo, Goseong and Jeju Islands. Static information of 21,049 ships was collected. There were ships that do not report length and breadth as well as ships that do not report ship type or have multiple ship types. Thus, we removed those ships from the dataset, then obtained 20,071 ships static data. Among them, the number of cargo is largest as 13,621 following by tankers with 5049, and others were 1407. To avoid biased training, we set 1407 per ship type, such as cargo, tanker, others, so that the total training datasets has 4221.

We used length and LBR, an abbreviation of length-to-breadth ratio, as features to training and testing. In length, the cargo and tanker can be discriminated against from others because they are not overlapped around the border on 105 m, as shown in Figure 4a. For breadth, the distribution is not distinctive from lengths, as shown in Figure 4b. Thus, the breadth cannot be trained to be a feature in terms of length. In LBR, on the other hand, cargo and tanker relatively distinguishable than length case, with exception of overlapped range 5.9 to 6.4, as shown in Figure 4c.
Figure 4. Length and length-to-breadth ratio (LBR) of text training data; (a) boxplot of length show that cargo and tanker can be distinguished from others. (b) Box plot of breadth shows similar distribution with length. (c) Boxplot of LBR represents cargo and tanker relatively well-identified except overlapped range 5.9 to 6.4. (d) Scatterplot between length and LBR shows that others widely scattered while the cargo and tanker are relatively concentrated.

LBR indicates the narrowness of the ships. The cargo, including the container and bulk, have a relatively narrow ship’s hull for high speed, the tanker is relatively wider than the cargo because they do not place importance on speed than cargo. Others have wide hulls compared to their low length of less than 100 m. The scatterplot shows how the scattered characteristics appear with their type. The features of cargo and tanker are relatively concentrated than others. They show that length alone cannot discriminate ship types. Meanwhile, LBR can be a supplement for classification.

3. Methodology

3.1. CNN Model

The CNN model consists of every single layer for image, output, flatten layer, fully connected layer and output score, and every repeated layer for convolution, activation and max-pooling, as shown in Figure 5. The convolution layer is generated from convolution between the image layer and weight filter, with one having size 3 (height) × 3 (breadth) × 1 (number of image channel) × 32 (number of output channel). The size 3 pixel is around 30 m in the Sentinel-1 IW GRD image; the nearest value to the mean 29.8 m and median 32.0 m of breadths of the static datasets. The 30 m can discriminate type others from cargo and tanker because the type others have 20.0 m for quantile 75%, as shown in Figure 4b. The values of the weight filter are assigned by the random sample and normal distribution.

ReLU, an abbreviation of rectified linear unit, is an activation function. It converts value less than zero into zero, but keeps over zero as it is as Equation (1). This prevents gradient vanishing, a symptom where gradient becomes zero in increasing layers, causing stopping training when training a neural network model using gradient-based learning and backpropagation.

\[
\text{ReLU}(x) = \begin{cases} 
0 & \text{for } x < 0, \\
 x & \text{for } x \geq 0 
\end{cases}
\] (1)

The max pooling resample activation map and return feature-abstracted image using a kernel with size 1 (number of input layer) × 2 (height) × 2 (breadth) × 1 (number of output value). The stride means how many pixels the kernel moves at once. We have stride size 2 (height) × 2 (breadth), then the max-pool kernel moves 2 pixels each in height and breadth direction at once.

The flatten layer from max-pooling layer 2 is multiplied with weight filter 3. The value of weight filter 3 is assigned by Xavier initialization [33]. By matrix product between flatten
layer and weight 3, it returns output scores, named logit, for three ship types. However, the logit itself is not effective enough to distinguish between the types. Thus, SoftMax function is used, a function that makes a low value to be lower and a high value to be higher, and their summation becomes value 1, as Equation (2). Among them, the one having the highest score chosen as predicted ship type and subsequently the predicted ship type is compared to the actual ship type.

\[
\text{Softmax}(X) = \frac{\exp(X)}{\sum \exp(X)}
\]  

(2)

where, \(X_{\text{new}} = X_{\text{old}} - \max(X_{\text{old}})\).

The difference of ship type scores between output and actual score can be compared by Cross-Entropy Error (CEE). This method, called SoftMaxWithLoss, as Equation (3), returns cost. For example, the output score has 0.72, 0.2 and 0.08 for cargo, tanker and others meanwhile actual scores are 1, 0 and 0. In this case, the cost is 0.328. To reduce the cost, the model used Adam optimizer, an abbreviation of adaptive moment estimation, which a combination method of momentum and RMSprop [34].

\[
\text{CEE}(Y, T) = \sum (T \times \log(Y))
\]  

(3)

where, \(Y_{\text{new}} = Y_{\text{old}} + 10^7\) (to prevent negative divergence).

CNN parameters are described in Table 1. Training epoch is achieved as the overall process of all data completed at a time, from the input image layer to the comparison between predict and actual ship type. Since we have 3780 training ship images, 1 epoch is made when 3780 data are trained. More training epoch results in higher test accuracy though, it gives a longer training duration.

![Figure 5. Overall Flow of CNN model. The phase from convolution layers to max-pooling layers is repeated. ReLu is used as an activation function to prevent gradient vanishing and max.](image)

| Parameter                  | Setting |
|----------------------------|---------|
| Training epoch             | 300     |
| Batch                      | 20      |
| Optimizer                  | Adam    |
| Learning rate              | 0.0001  |
| Cost function              | CEE     |

The learning rate is how much change the weight to minimize cost at once. If the learning rate too low, the work to find the best weight is finished due to the epoch. If too high, the model cannot find the weight because subtracted weight repeatedly surpassed the best weight.

Batch, an abbreviation of mini-batch, means the number of data that will be processed at once. It is faster to process 20 data 189 times rather than 1 data 3780 times. Larger batch
size has faster training though, and it makes test accuracy lower. In most cases, less than 32 is proper [35].

After completing 300 training epochs, CNN finally predicts ship type by finding the highest one among the output scores.

3.2. KNN

KNN is an algorithm that finds the \( k \) elements closest to the input in a specific space and classifies them as more matches [31]. As described in Section 2.2, 4221 pairs of length and LBR were used as training features to classify ship types such as cargo, tanker, and others. KNN predicts ship type with the following steps.

Integer \( k \) should be assigned first. According to dataset distribution, the optimized \( k \) is determined so that we iterated integer 1 to 20 and found 6 to be the best \( k \) that results in the highest accuracy. The length and breadth pass through min-max normalization and z-score standardizations, as shown in Equations (4) and (5). This is because when a pair of features is inputted, the KNN searches the nearest \( k \)-pair of features using Euclidean distance on the same scale.

\[
\text{Min-Max normalization} \ (X) = \frac{(X - \text{min}(X))}{(\text{max}(X) - \text{min}(X))} \tag{4}
\]

\[
\text{z-score standardization} \ (X) = \frac{(X - \mu)}{\sigma} \tag{5}
\]

where, \( \mu \) is mean \((X)\) and \( \sigma \) is StdDev \((X)\).

Through the proportional probability of three ship types, a ship type is predicted. For instance, \( k \) is 3 and KNN found three pairs of features. If cargo, tanker and others have 1, 2 and 0, respectively, KNN predicts the ship type to be a tanker.

3.3. Confusion of CNN and KNN Probability

In this paper, we propose a method to improve the prediction ability by using CNN and KNN together. Image-based probability \((P_I)\) is from a fully connected layer after 300 epochs of training as described in Section 3.1. Instead of the output score after SoftMax, the logits in a fully connected layer are normalized then converted into proportional probability for three ship types. On the other hand, text-based probability \((P_T)\) is from KNN, as shown in Section 3.2.

There are four distinctive method proposals to determine the ship type as shown in Figures 6 and 7. The first method, named Max \((I, T)\), is to get a label having the highest value between \(\text{Maximum} \ P_I\ (MP_I)\) and \(\text{Maximum} \ P_T\ (MP_T)\). The second method, named Ave \((I, T)\) is the mean value between \(P_I\) and \(P_T\) for every ship type. This method gets an average probability for cargo, tanker, and others \((ACTO)\), respectively, from \(P_I\) and \(P_T\), and then the label having maximum average value among the three ship types is chosen as the final type.

**Figure 6.** Combination methods to choose type using maximum probability and average probability. (a) Max \((I, T)\) method choose the label having maximum probability; (b) average \((I, T)\) method choose the highest one among \(ACTO\). Average probability is from the average of \(P_I\) and \(P_T\) each for cargo, tanker and others.
The third and fourth methods use threshold for $P_T$ first. As threshold value for $P_T$, we used 0.83 which most frequent probability between 0.5 and 1.0 in KNN classification result. If a ship has $P_T$ equal to or more than 0.83 for ship types, both the third and fourth methods, named Cond_Max (I, T) and Cond_Std (I, T), the ship type is predicted by the value. If not, it compares the Max $P_I$ and Max $P_T$, then gets the highest one as Figure 7a or it compares the standard deviation of $P_I$ and $P_T$, as in Figure 7b. The overall process is described in Figure 8.
4. Results

For the evaluation of the prediction result with test datasets, we used accuracy, precision, recall and f1-score defined by Equations (6)–(9). The four evaluation parameters are computed using true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Among them, TP and FN are correct predictions meanwhile FP and FN are incorrect.

Precision is the ratio of what the model classifies as true that is actually true, whereby the actual cargo among the predictions is cargo. Recall is the percentage of what the model predicts as true among what is actually true, whereby the prediction is cargo among what is actually cargo. The accuracy represents the cases where the predictions are true as well as false. However, the accuracy is affected by the biased number of labels. The F1-score evaluates the model’s performance by harmonic mean of precision and recall. Precision, recall and f1-scores are calculated by macro average, where the parameters are calculated for each label, then averaged.

\[
\text{Precision} \% = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100
\]

\[
\text{Recall} \% = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100
\]

\[
\text{Accuracy} \% = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \times 100
\]

\[
\text{f1 score} \% = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Ship type classification is made by five kinds of methods, as shown in Table 2. The first is to use images alone as training and testing datasets. The rest of the four methods are the combination methods, as described in Section 3.3.

**Table 2.** Result of classification according to image type and classification method.

| Method    | Image Type | Precision (%) | Recall (%) | Accuracy (%) | F1-Score (%) |
|-----------|------------|---------------|------------|--------------|--------------|
| Image alone | VH         | 85.6          | 86.3       | 85.6         | 85.0         |
|           | VV         | 83.3          | 83.3       | 83.3         | 83.2         |
|           | MaxVHV     | 82.2          | 83.5       | 82.2         | 81.2         |
|           | MinVHV     | 84.4          | 85.2       | 84.4         | 84.1         |
| Max (I, T) | VH         | 90.0          | 90.3       | 90.0         | 90.0         |
|           | VV         | 92.2          | 92.5       | 92.2         | 92.2         |
|           | MaxVHV     | 91.1          | 91.3       | 91.1         | 91.0         |
|           | MinVHV     | 92.2          | 92.3       | 92.2         | 92.1         |
| Ave (I, T) | VH         | 93.3          | 93.9       | 93.3         | 93.1         |
|           | VV         | 94.4          | 94.9       | 94.4         | 94.3         |
|           | MaxVHV     | 93.3          | 93.5       | 93.3         | 93.2         |
|           | MinVHV     | 94.4          | 94.9       | 92.2         | 94.3         |
| Cond.Std (I, T) | VH     | 91.1          | 91.3       | 91.1         | 91.0         |
|            | VV         | 93.3          | 93.5       | 93.3         | 93.3         |
|            | MaxVHV     | 90.0          | 90.2       | 90.0         | 89.9         |
|            | MinVHV     | 92.2          | 92.3       | 92.2         | 92.1         |
| Cond_MAX (I, T) | VH    | 91.1          | 91.3       | 91.1         | 91.0         |
|             | VV         | 92.2          | 92.5       | 92.2         | 92.2         |
|             | MaxVHV     | 91.1          | 91.3       | 91.1         | 91.0         |
|             | MinVHV     | 92.2          | 92.3       | 92.2         | 92.1         |

Four parameters, including precision, recall, accuracy, and f1-score in images, are less than 90.0%. Meanwhile, the proposed four methods are equal to or over 90.0% in most cases. This shows that combination methods are effective to increase classification performance.
In the Image alone method, VH and VV have 85.0% and 83.2% in f1-score. It shows that VH gives high contribution than VV, as referred to introduction part. MaxVHV and MinVHV do not perform better than VH and VV methods. Thus, the newly made images are not always effective.

Among combination methods, the f1-score was higher in the order of Ave (I, T), Cond_Std (I, T), Cond_Max (I, T) and Max (I, T). Thus, Ave (I, T) is the most preferable as a combination method.

Among the results in Table 2, we concluded that VV and Ave (I, T) is effective in CNN and KNN combination classification, showing its highest accuracy to be 94.4%.

Figure 9 shows the confusion matrix for the image alone and combination methods with VH polarization. The test result of the image alone shows that the model was confused with cargo and tanker. It may be a result from the similar shapes of cargo and tanker in their images, although their brightness distribution against the ships’ hull is different. Meanwhile, every combination method increased the classification ability of the cargo and tanker.

VGG19 and ResNet50, well-known classification models [35,36], are chosen to compare the performance of the proposed KNN and CNN combination methods. VGG19 consists of 16 convolution layers, 3 fully connected layers, 5 max-pooling layers and 1 SoftMax layer [36]. ResNet50 has 48 convolution layers, 1 max-pooling layer, and 1 average-pooling layer, and it uses bottleneck architecture, which prevents the vanishing gradient problem [14]. Both models could be operated with transfer learning (TL) in which a pre-trained CNN model with large datasets, is often adopted to improve the classification performance with small datasets [37].

Training and test datasets, batch size, and epoch are identical with the CNN model we presented in Section 3.1. For VGG19 and ResNet50, both were experimented on in the cases with and without TL, as in Table 3. VGG19 shows the lowest performance accuracy of 33.3% in VH and VV, meanwhile ResNet50 has over the moderate performance of more than or equal to 70%. MaxVHV and MinVHV commonly have low performances in every image type. Transfer learning improved their classification ability, increasing the accuracy up to 45.6% and 44.5% of VGG19 and ResNet50, respectively. The Ave (I, T) proposed in this study shows the accuracy of 93.1% in VH meanwhile VGG19-TL and ResNet50-TL show 78.9% and 81.1%, respectively.
Table 3. Result of classification using VGG19 and ResNet50.

| Model   | Image Type | Precision (%) | Recall (%) | Accuracy (%) | F1-Score (%) |
|---------|------------|---------------|------------|--------------|--------------|
| VGG19   | VH         | 33.3          | 11.1       | 33.3         | 16.7         |
|         | VV         | 33.3          | 11.1       | 33.3         | 16.7         |
|         | MaxVHV    | 33.3          | 11.1       | 33.3         | 16.7         |
|         | MinVHV    | 33.3          | 11.1       | 33.3         | 16.7         |
| VGG19-TL| VH         | 78.9          | 78.8       | 78.9         | 77.9         |
|         | VV         | 78.9          | 78.5       | 78.9         | 78.4         |
|         | MaxVHV    | 76.7          | 76.4       | 76.7         | 75.6         |
|         | MinVHV    | 74.4          | 73.4       | 74.4         | 73.1         |
| ResNet50| VH         | 77.8          | 77.5       | 77.8         | 76.4         |
|         | VV         | 70.0          | 73.5       | 70.0         | 68.3         |
|         | MaxVHV    | 33.3          | 11.1       | 33.3         | 16.7         |
|         | MinVHV    | 33.3          | 11.1       | 33.3         | 16.7         |
| ResNet50-TL| VH | 81.1          | 81.2       | 81.1         | 80.4         |
|         | VV         | 81.1          | 80.9       | 81.1         | 80.3         |
|         | MaxVHV    | 80.0          | 79.7       | 80.0         | 79.0         |
|         | MinVHV    | 77.8          | 77.2       | 77.8         | 77.0         |

5. Discussion

Image types VH, VV, maxVHV and minVHV show different classification result in Table 2. Every image type has high frequencies near a probability of 0.5, as shown in Figure 10. If the distribution is clustered nearby the average, it means that the model has high ambiguity to determine ship types. Meanwhile, if the probabilities are distributed toward both lower and higher, it represents that the model more certain to determine ship types as positive or negative.

![Figure 10. Distribution of ship type probability according to image types.](image)

*P_I* and *P_T* are abbreviation of image- and text-based probability. (a–d) show *P_I* for four type images and *P_T*. *P_T* has more frequency on probability 1 and 0 than *P_I* meanwhile *P_I* has more probability 0.33, 0.5 and 0.67.

*P_T* has more confidence than *P_I* as shown in Figure 9. *P_T* has discrete probability such as 0, 0.17, 0.33, 0.50, 0.67, 0.83 and 1. These are the results of the best k value we obtained in Section 3.2.

Among combination methods Ave (I, T) was shown as the most effective. Figure 11 shows image alone and Ave (I, T) using VH polarization. Ave (I, T) reduced the frequency
of center-clustered probabilities and increased side probability which are under 0.33 or over 0.67. It illustrates that how the method reduced ambiguity in determining ship types.

![Figure 11](image.png)

**Figure 11.** Image alone with VH and Ave (I, T) with VH. In an image-alone case, the probabilities are center-clustered. Meanwhile, Ave (I, T) reduced the center-clustered probabilities and distributed them to less than 0.33 and more than 0.67.

### 6. Conclusions

This study proposed to combine CNN and KNN to improve ship type classification performance. Ships’ length and length-to-breadth ratio (LBR) were selected as the features of KNN. Because each CNN and KNN use different features as image and text, the probabilities of ship type from both models were utilized as a common parameter for combination methods. The proposed four combination methods showed the enhanced classification performance rather than the case CNN trained with ship images only.

In the future, some improvements are still required. The number of training datasets needs to be increased and new types of data such as minVHVV and maxVHVV will be prepared. Additionally, an optimized method combining CNN and KNN will be investigated including a ship dimension extraction from SAR data.

**Author Contributions:** Conceptualization, C.-S.Y.; methodology, H.-K.J. and C.-S.Y.; software, H.-K.J.; validation, H.-K.J.; formal analysis, H.-K.J.; investigation, H.-K.J.; resources, C.-S.Y.; data curation, H.-K.J.; writing—original draft preparation, H.-K.J.; writing—review and editing, H.-K.J.; visualization, H.-K.J.; supervision, C.-S.Y.; project administration, C.-S.Y.; funding acquisition, C.-S.Y. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research is a part of the projects entitled “Development of satellite based system on monitoring and predicting ship distribution in the contiguous zone”, funded by the Korea Coast Guard, and “Establishment of the ocean research station in the jurisdiction zone and convergence research”, funded by the Ministry of Oceans and Fisheries, Korea.

**Data Availability Statement:** Restrictions apply to the availability of these data. OpenSARship was obtained from OpenSAR website and are available at URL (https://opensar.sjtu.edu.cn accessed on 1 May 2021) with the permission of OpenSAR.

**Acknowledgments:** The authors would like to thank Seungryong Kim who gave help in constructing CNN model and all of the editors/anonymous reviewers for their careful reading and insightful remarks.

**Conflicts of Interest:** The authors declare no conflict of interest.
28. Eriksen, T.; Høye, G.; Narheim, B.; Meland, B.J. Maritime traffic monitoring using a space-based AIS receiver. *Acta Astronaut.* 2016, **58**, 537–549. [CrossRef]

29. Hong, D.-B.; Yang, C.-S.; Kim, T.-H. Investigation of Passing Ships in Inaccessible Areas Using Satellite-based Automatic Identification System (S-AIS) Data. *Korean J. Remote Sens.* 2018, **34**, 579–590.

30. Parsa, A.; Hansen, N.H. Comparison of Vertically and Horizontally Polarized Radar Antennas for Target Detection in Sea Clutter. In Proceedings of the 2012 IEEE Radar Conference, Atlanta, GA, USA, 7–11 May 2012; pp. 653–658.

31. Fix, E.; Hodges, J.L. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *Int. Stat. Rev.* 1989, **57**, 238–247.

32. OpenSAR. Available online: https://opensar.sjtu.edu.cn (accessed on 11 May 2021).

33. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *J. Mach. Learn. Res.* 2010, **9**, 249–256.

34. Kingma, P.; Ba, J.L. Adam: A Method for Stochastic Optimization. *arXiv* 2017, arXiv:1412.6980.

35. Dominic Master, M.; Luschi, C. Revisiting Small Batch Training for Deep Neural Networks. *arXiv* 2018, arXiv:1804.07612.

36. Simonyan, K.; Andrew, Z. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, arXiv:1409.1556.

37. Bozinovski, S.; Fulgosi, A. The influence of pattern similarity and transfer learning upon the training of a base perceptron B2. (original in Croatian). In Proceedings of the Symposium Informatica 3-121-5, Bled, Slovenia, 1976. Available online: http://www.informatica.si/ojs-2.4.3/index.php/informatica/article/viewFile/2828/1433 (accessed on 13 May 2021).