Mathematical Biosciences and Engineering

MBE, 20(10): 18545–18565.
DOI: 10.3934/mbe.2023823
Received: 21 July 2023
Revised: 02 September 2023
Accepted: 11 September 2023
Published: 27 September 2023

http://www.aimspress.com/journal/MBE

Research article

Maritime ship recognition based on convolutional neural network and linear weighted decision fusion for multimodal images

Yongmei Ren[1], Xiaohu Wang[2,*] and Jie Yang[3]

[1] School of Electrical and Information Engineering, Hunan Institute of Technology, Hengyang 421002, China
[2] College of Intelligent Manufacturing and Mechanical Engineering, Hunan Institute of Technology, Hengyang 421002, China
[3] Hubei Key Laboratory of Broadband Wireless Communication and Sensor Networks, School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China

* **Correspondence:** Email: 2003000285@hnit.edu.cn.

Abstract: Ship images are easily affected by light, weather, sea state, and other factors, making maritime ship recognition a highly challenging task. To address the low accuracy of ship recognition in visible images, we propose a maritime ship recognition method based on the convolutional neural network (CNN) and linear weighted decision fusion for multimodal images. First, a dual CNN is proposed to learn the effective classification features of multimodal images (i.e., visible and infrared images) of the ship target. Then, the probability value of the input multimodal images is obtained using the softmax function at the output layer. Finally, the probability value is processed by linear weighted decision fusion method to perform maritime ship recognition. Experimental results on publicly available visible and infrared spectrum dataset and RGB-NIR dataset show that the recognition accuracy of the proposed method reaches 0.936 and 0.818, respectively, and it achieves a promising recognition effect compared with the single-source sensor image recognition method and other existing recognition methods.

Keywords: maritime ship recognition; convolutional neural network; linear weighted decision fusion; visible images; infrared images

1. Introduction

Ship recognition has broad application prospects in maritime safety, traffic control, and precision-guided weapons [1]. The automatic control of sea surface can realize the visual automatic ship recognition, which has attracted extensive research attention. Synthetic aperture radar (SAR) images can be acquired at all day/night and are not limited by light, weather and climate conditions. However, SAR images require expensive acquisition equipment. They are mainly used in the military field. Visible images can directly display the detailed information of ship targets with a high resolution, but the imaging effect is poor in extreme weather, such as rain and fog and night conditions. Infrared images have strong penetration ability and can distinguish a target from its background based on radiation differences. They can be acquired in any weather and have a clear target contour, but they have low resolution. Therefore, the ship image information obtained by a variety of sensors is complementary, and the fusion of multimodal images for ship recognition has greater advantages than using single-source sensor image for recognition. This subject has gradually become a research hotspot in the field of computer vision.

Maritime ship recognition methods mostly include traditional recognition and convolutional neural network (CNN)-based methods. Traditional maritime ship recognition methods adopt handcrafted features for recognition, including local binary pat-tern (LBP) [2], scale-invariant feature transformation [3] and histogram of oriented gradients (HOG) [4]. Handcrafted features rely on expert knowledge and have poor generalization ability. Therefore, the recognition ability of traditional maritime ship recognition methods is limited.

Deep learning technology has significantly improved the performance of computer vision tasks and has been widely used in image recognition [5,6] and pedestrian detection [7]. Zhang et al. [8] fused HOG features with improved CNN to improve the ship recognition performance of SAR images; this method had a recognition accuracy that was 7.64% improved compared with that of the CNN method. Xu et al. [9] proposed a ship recognition method by combining CNN and attention mechanism. In [10], two CNNs were designed and transfer learning was used to identify ships; the effectiveness of this method was verified on the ship dataset with a comparatively small number of samples. Li et al. [11] proposed a ship recognition method based on improved Faster R-CNN for SAR images; this method introduced transfer learning and feature aggregation to improve the average accuracy. Wang et al. [12] presented a ship recognition method based on single shot multibox detector (SSD), and transfer learning is used to solve the problem of insufficient training samples; this method achieved good recognition results in Chinese Gaofen-3 SAR images. Wang et al. [13] proposed a ship recognition method based on SSD and transfer learning in complex background, which improved the recognition performance. In [14], a classification method based on deep transfer learning is proposed to solve the SAR image classification problem with a few labeled data. Ganesh et al. [15] designed a real-time video processing method for ship detection based on transfer learning, and transfer learning technology is used to train the model. Shi et al. [16] fused CNN with multifeatures to classify ships. This method showed better classification performance than the single feature-based methods, but the algorithm complexity was higher. Mishra et al. [17] used the transfer learning method based on pre-trained VGG16 to classify four types of ships. However, the dataset used was small and the algorithm lacked generalization ability. Wang et al. [18] proposed a ship recognition method based on multi-scale feature attention and adaptive weighted classifier, which improved the performance of SAR ship recognition. Ucar et al. [19] proposed a ship classification method based on deep cascade network. AlexNet and

VGG16 networks were used to learn the deep features and they were integrated in the full connection layer. Mutual information feature selection method was applied to construct the deep feature set, and finally, SVM classifier was used for classification. Aziz et al. [20] proposed visible and infrared spectrum recognition method based on multimodal CNN, but the ship recognition accuracy had to be improved. In [21], a ship recognition method based on CNN is proposed. Transfer learning training model is used to avoid overfitting problem and improve the recognition performance on a small dataset. Qiu et al. [22] proposed a two-band decision fusion ship recognition method by combining multilayer convolution features and posterior probability weighting, but the recognition accuracy was 89.7%, which needed improvement. In [23], the improved SqueezeNet was proposed to extract features, the shallow and deep features were cascaded, and the Adam optimizer was improved to increase the ship recognition accuracy. Du et al. [24] proposed a ship detection and recognition fusion classifier based on CNN, and the classification accuracy reached 84.7% on the self-built visible ship image dataset. Zhang et al. [25] proposed a fine-grained ship recognition method based on Inception and VGG16; this method used AM-Softmax to obtain predicted labels and verified the validity of the method on the self-built dataset. Huang et al. [26] proposed a new ship classification and detection method by combing CNN and Swin Transformer. Self-attention mechanism was introduced into Transformer to achieve good recognition effect. Wang et al. [27] proposed a lightweight improved GhostNet-50 to identify a self-made ship dataset. Compared with GhostNet-50, the new model was compressed by 46.67%. However, the recognition capability of the model had to be improved.

In the preceding studies, most of the methods only adopted SAR images, visible images and infrared images to identify ships, and the complementary information between multimodal images was less considered. Maritime ship recognition accuracy needed to be improved. In addition, the maritime ship recognition method for multimodal images encountered problems such as low concatenated feature fusion quality and high algorithm complexity. Based on the recognition results of various classifiers, decision fusion gave the final decision results in the fusion center according to certain fusion rules. Decision fusion had less error information than the single classifier in recognition [28], with successful adoption in pedestrian detection [29] and action recognition [30] based on multimodal images. Considering the low recognition accuracy in maritime ship recognition for visible images, we proposed a maritime ship recognition method in multimodal images with CNN and linear weighted decision fusion. First, the dual CNN composed of visible and infrared subnetworks had been used to extract and learn the effective classification features of multimodal images. Then, the probability value classified by softmax function was processed through linear weighted decision fusion to maximize the complementary advantages of the effective classification features in the dual CNN and the probability value of the multimodal images, and to improve the ship recognition accuracy.

This study has the following major contributions:

• An improved CNN with fewer convolutional layers and kernels is proposed. The improved CNN can avoid the overfitting phenomenon caused by the few labeled samples.

• Based on the improved CNN, a dual CNN is proposed as a feature extractor to extract effective classification features of multimodal images to ensure that the extracted features contain more semantic features and distinctive information.

• The linear weighted decision fusion model is constructed to give appropriate weights to the visible and infrared subnetworks, and process the probability value classified by the softmax function. The model can effectively use the complementary advantages of effective classification features within multimodal images and improve the recognition ability of maritime ship recognition method.

The rest of this paper is organized as follows. Section 2 introduces the proposed maritime ship recognition model. Section 3 describes experimental environment and evaluation metrics, the VAIS dataset [31], RGB-NIR dataset [32] and reports the experimental results. Section 4 summarizes the research results and future research directions.

2. Recognition method based on dual CNN and linear weighted decision fusion

Visible images are easily affected by light, which sometimes leads to unclear image details and causes the misrecognition of ship images. Infrared images can be obtained in any weather, but most of the images have low resolution and cannot reflect the color information of the target. Therefore, using only a single source image for ship recognition is a challenging task. The maritime ship recognition method based on deep CNN can acquire abstract feature representation from ship images. Therefore, to solve the problem of low ship recognition accuracy in visible images, we propose a maritime ship recognition method by combining CNN and linear weighted decision fusion for multimodal images. Figure 1 shows the flow chart of this method, which mostly includes a preprocessing module, a feature extraction module and a decision fusion recognition module. In the training phase, the multimodal images are preprocessed (the images are resized in Section 3.1). Then, the dual CNN composed of a visible subnetwork and an infrared subnetwork is adopted to learn the effective classification features of the preprocessed multimodal images. We use the softmax function to process the extracted effective classification features to obtain the predicted ship category labels in the output layer. The errors between the true ship category labels and the predicted ship category labels are calculated. The back propagation algorithm is used to iteratively update the weight and bias until the errors are minimized. Finally, the optimal training model of the dual CNN is obtained and saved. In the testing phase, the preprocessed multimodal images of the same ship target are input into the visible subnetwork and the infrared subnetwork, respectively, to extract the effective classification features. The optimal model is called to test them. After that the probability value classified by the softmax function is obtained. Then, the linear weighted decision fusion method is used to process the probability value, and the final recognition results of maritime ship images are obtained.

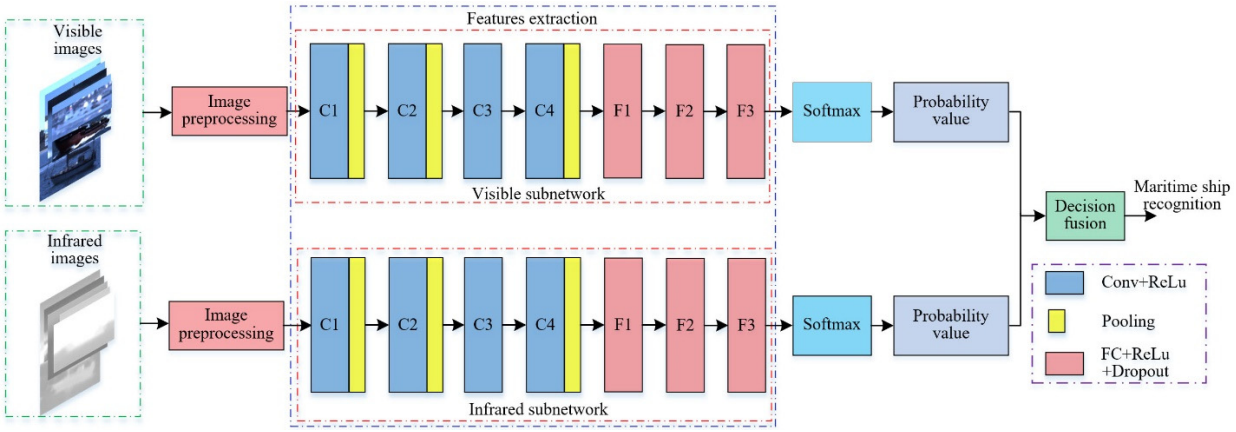


Figure 1. Framework of recognition method based on dual CNN and linear weighted decision fusion.

2.1. Convolutional feature extraction module

CNN combines feature learning with classifier training, and it can mine local features of input data [33]. CNN can realize most neuron weights sharing through local perception, parameter sharing, downsampling and other methods, so the number of parameters in the network can be reduced. Network architecture is important for the correct recognition of ship images. The parallel visible subnetwork and infrared subnetwork are selected to form a dual CNN, as shown in Figure 2. The dual CNN can learn the effective classification features of multimodal images. The visible and infrared subnetworks are two identical improved CNNs. The classic AlexNet [34] consists of five convolutional layers. Using AlexNet to train the ship dataset with few samples causes overfitting. In the visible subnetwork, the number of convolution layers is four, which not only learn the abstract features of the visible images but also reduce the computational complexity as far as possible and improve the overfitting phenomenon caused by the small dataset. A fully connected layer is added to obtain more semantic features on the multimodal ship images and the number of neurons is 2048. The rectified linear unit (ReLU) function is more consistent with sparse activation of biological neurons, which can accelerate network convergence and reduce gradient disappearance in the training network. Therefore, we use the ReLU function in the convolutional and fully connected layers. The maximum pooling method is used to reserve more texture information of the multimodal images and reduce the dimensions of the convolution results for the upper layer.

We use dropout in fully connected layers to avoid network overfitting. In the output layer we use the softmax function to predict class labels for multimodal images. Table 1 lists the parameters of the improved CNN. The softmax output node is equal to n, which corresponds to the number of multimodal images classes contained in the VAIS dataset and RGB-NIR dataset (refer Section 3.1). Figure 3 shows the improved CNN structure. Conv1 in the figure represents the first convolution layer, Pool1 in the figure represents the maximum of the first pooling layer. FC1 represents the first fully connected layer. Padding means adding 0 to the outer layer. If the value is two, then it means that the outer layer is expanded with two circles of 0.

In this study, training samples are randomly cropped into 227 × 227 to increase training samples and ensure the diversity of samples. The test samples are center cropped. Data enhancement method such as random horizontal flip is also used to improve the generalization ability of the proposed method.

Table 1. Parameters of improved CNN.

Layer	Filter Number	Kernel Size/Stride	Padding
Conv1	64	11 × 11/4	2
Pooling1	–	3 × 3/2	–
Conv2	192	5 × 5/1	2
Pooling2	–	3 × 3/2	–
Conv3	384	3 × 3/1	1
Conv4	256	3 × 3/1	1
Pooling3	–	3 × 3/2	–
FC1	–	4096	–
FC2	–	4096	–
FC3	–	2048	–
Softmax	–	*n*	–

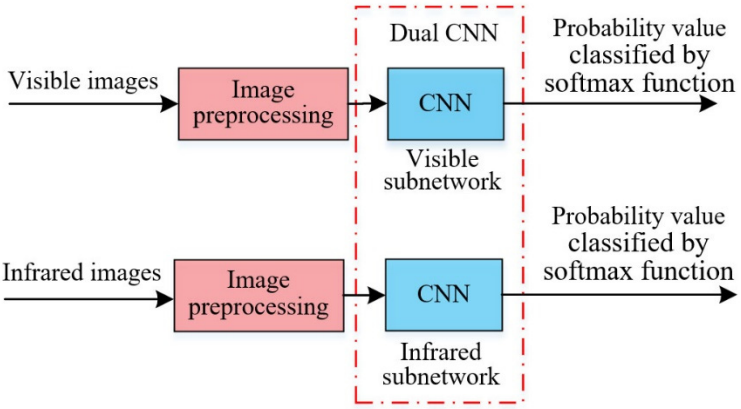


Figure 2. Dual CNN.

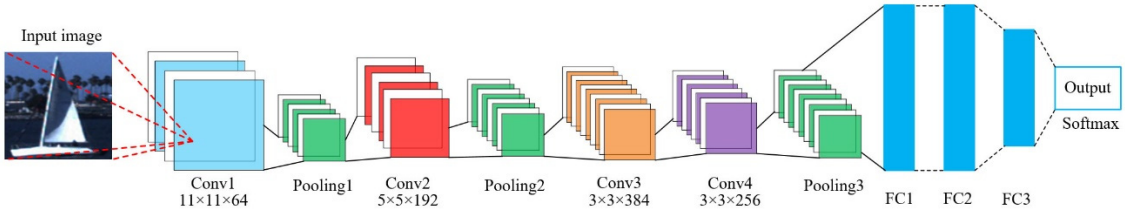


Figure 3. Network structure of improved CNN.

2.2. Linear weighted decision fusion method

Decision fusion can have better recognition results by processing the output results of different classifiers. The weight of each classifier affects the recognition effect of decision fusion. Therefore, when the output result of classifiers is probabilistic value, each classifier is assigned an appropriate weight to represent the contribution of various features in the recognition process. The result of decision fusion can be obtained through linear weighted summation of the output probability values of different classifiers. To improve the performance of maritime ship recognition using single source images, we construct a linear weighted decision fusion model to process the probability value classified by the softmax function. It can combine the effective information of multimodal images, obtain more accurate recognition results, and maintain the feature dimension without increasing the parameters of the model. The specific linear weighted decision fusion method is as follows:

The probability value classified by softmax function can be denoted as

$$\boldsymbol{P}(x)=\begin{bmatrix} p_{11}(x) & p_{12}(x)\cdots p_{1i}(x) \\ p_{21}(x) & p_{22}(x)\cdots p_{2i}(x) \end{bmatrix}_{2\times i} \tag{1}$$

where the first row of the matrix represents the input sample probability value classified by the softmax function of the visible subnetwork, and the second row represents the input sample probability value classified by the softmax function of the infrared subnetwork. x indicates input sample and i indicates the number of categories in the dataset. The label of the column with the highest probability

in each row is the prediction category for the sample by the softmax function of each subnetwork. Based on the assumption that α and β are the weights of the probability value classified by the softmax function of the visible and infrared subnetworks, respectively, the new probability value output matrix can be defined as

$$\boldsymbol{P}'(x)=\begin{bmatrix}\alpha p_{11}(x) & \alpha p_{12}(x)\cdots\alpha p_{1i}(x)\\ \beta p_{21}(x) & \beta p_{22}(x)\cdots\beta p_{2i}(x)\end{bmatrix}_{2\times i} \tag{2}$$

where $\beta=1-\alpha$, $0<\alpha<1$, $0<\beta<1$. $\alpha=0$ means only ships with infrared images have been identified and $\alpha=1$ means only identify ships with visible images have been identified.

The matrix of Eq (2) is weighted and summed according to the column. The label of the maximum value is the recognition result of ship images after linear weighted decision fusion processing, which can be expressed as

$$\text{label}(x)=\underset{j=1,2,\cdots,i}{\arg\max}\left[\alpha p_{1j}(x)+\beta p_{2j}(x)\right] \tag{3}$$

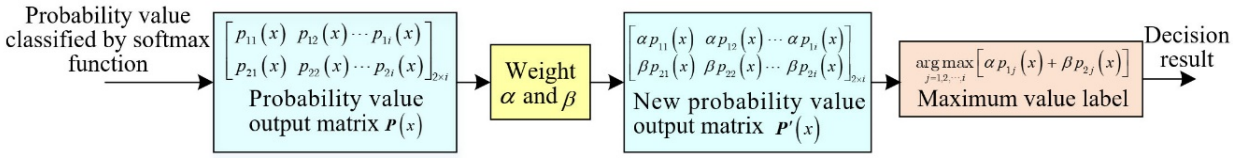
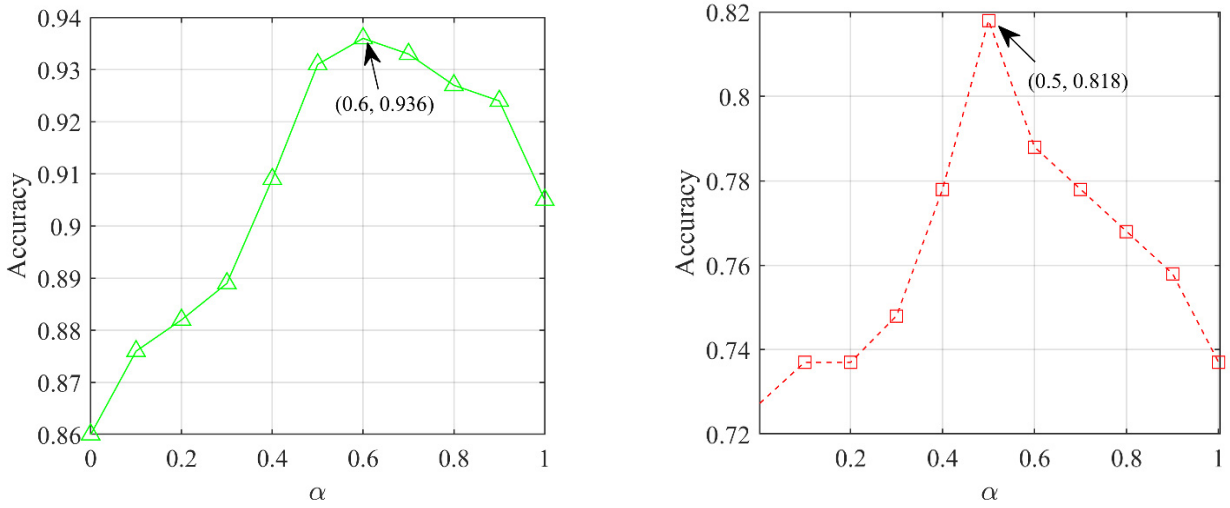


Figure 4. Flow chart of linear weighted decision fusion model.

Figure 4 shows the flow chart of the linear weighted decision fusion model obtained from the preceding calculation process.

To explore the role of parameter α in decision fusion, experiments were conducted on the VAIS and RGB-NIR datasets (refer Section 3.1) to find the α value that fits the dataset. Figure 5(a) shows the recognition accuracy in different α of the VAIS dataset. With the constant increase of α value, ship recognition accuracy also shows an improvement within a certain range. When α is 0.6, the ship recognition accuracy is the highest, reaching 0.936. This result shows that the visible images have a great influence on the ship recognition results. With the further increase of α, the ship recognition accuracy gradually decreases. The reason is that when the proportion of the ship recognition result for the infrared images decreases, there is a problem of misrecognition by using only the visible images for ship recognition. Therefore, for VAIS dataset, α is set to 0.6 in this study to obtain the optimal ship recognition performance.

The recognition accuracy in different α of the RGB-NIR dataset is shown in Figure 5(b). When α is 0.5, the recognition accuracy is the highest, reaching 0.818. The preceding results show that the linear weighted decision fusion method assigns appropriate weights to the visible and infrared subnetworks according to the different recognition results of the input multimodal images, and it solves the problem of misrecognition for the visible images. Therefore, it can increase the recognition accuracy.



(a) VAIS dataset (b) RGB-NIR dataset

Figure 5. Recognition accuracy in different α.

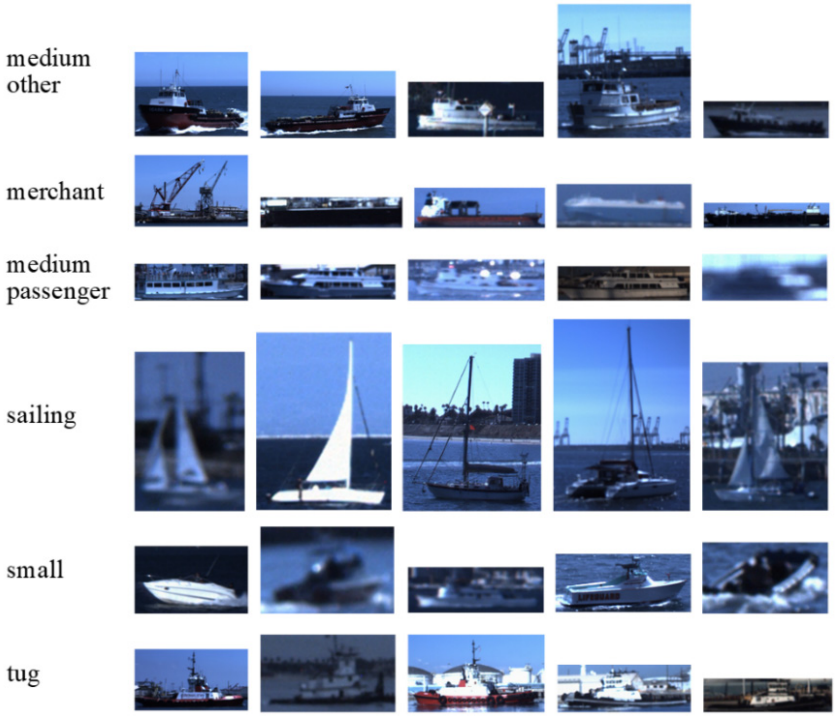
3. Experiments and analysis

3.1. Experimental dataset

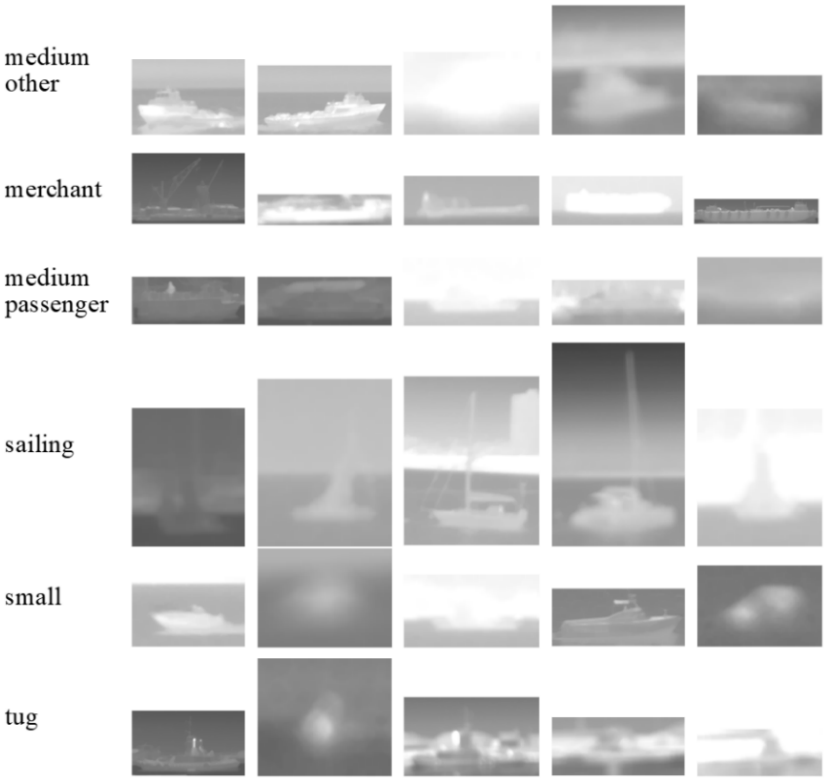
Experiments were conducted on two datasets to validate the proposed method.

The first public multimodal images dataset is VAIS [23], which consists of 2865 images and 1088 paired multimodal images. The dataset includes six categories: medium "other" ships, merchant ships, medium passenger ships, sailing ships, small boats and tugboats. Figure 6 shows the samples in the VAIS dataset. The numbers of each category are 138, 146, 117, 284, 353 and 50, respectively. The training set was obtained by random selection. The training samples consist of 539 pairs, in which the number of medium-other, merchant, medium-passenger, sailing, small and tug is 62, 83, 58, 148, 158 and 30, respectively. The remaining 549 pairs multimodal images are used as test samples. We resize multimodal images to 256 × 256 by adopting bicubic interpolation.

The second experimental dataset is RGB-NIR [24], which contains 477 paired scene images. The dataset includes nine categories: country, field, forest, indoor, mountain, old building, street, urban, and water, as shown in Figure 7. Each category has a visible image on the left and an infrared image on the right. Although the dataset is small, it contains categories that interfere with each other, such as country and field, street and city. The numbers of each category are 52, 51, 53, 56, 55, 51, 50, 58 and 51, respectively. Eleven pairs multimodal images were selected from each class randomly as test samples, and the remaining multimodal images as training samples.



(a) Visible images



(b) Infrared images

Figure 6. Samples in VAIS dataset.

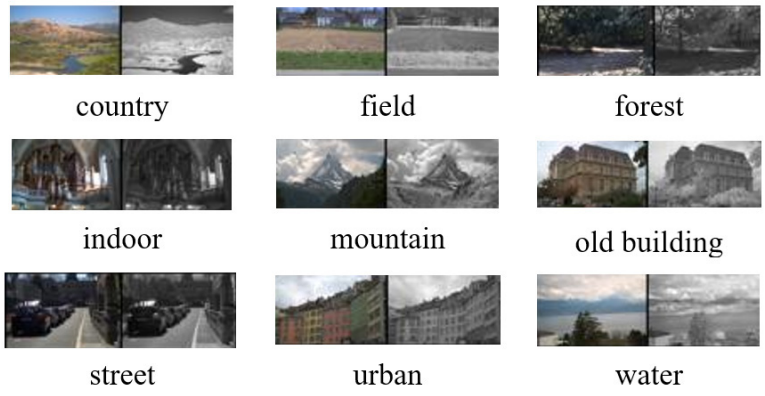


Figure 7. Samples in RGB-NIR dataset.

3.2. Experimental platform and parameter setting

The hardware environment of the experimental platform is Inter(R) Core(TM) i9-7980XE @ 2.6 GHz processor, 32 GB memory and GPU of NVIDIA TITAN Xp Pascal. The software environment is Python language and Pytorch framework.

Experimental parameter settings for the proposed method: the learning rate of the visible and infrared subnetworks is set as 0.001 in accordance with stochastic gradient descent method. The dropout is 0.5. The batch size is 32. The momentum parameter is 0.9, and the weight coefficient is 0.0001. The difference is that for the VAIS dataset, the learning epochs of the visible and infrared subnetworks are equal to 400 and 395, respectively. For the RGB-NIR dataset, the learning epochs of the visible and infrared subnetworks are 300.

3.3. Evaluation metrics

The recognition accuracy, the number of misrecognition samples, precision, recall, F1-score and feature extraction time per image are considered evaluation metrics of maritime ship recognition results.

Recognition accuracy represents the ratio of correctly identified multimodal images to the total number of multimodal images. The recognition accuracy can be defined as:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

where TP denotes the number of true positives, FP denotes the number of false positives, FN denotes the number of false negatives and TN denotes the number of true negatives.

The number of misrecognized multimodal images is obtained by calculating the product of the total number of multimodal images and the error rate.

Precision indicates the percentage of true positives in the predicted positives. It can be defined as:

$$P = \frac{TP}{TP + FP} \tag{5}$$

Recall represents the percentage of samples that are predicted to be positive in the true positives. It can be expressed as:

$$R = \frac{TP}{TP + FN} \tag{6}$$

F1-score is a metric of comprehensive performance. The value of F1-score ranges from 0 to 1. 1 represents the best output result of the model and 0 represents the worst output result of the model. It can be evaluated as:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{7}$$

The confusion matrix is a visual tool. Each column of the confusion matrix denotes the predicted category, total number of each column denotes the number of multimodal images predicted for the corresponding category, each row denotes the real category, and total number of each row denotes the number of real multimodal images for the corresponding category.

3.4. Recognition results and Analysis

The effectiveness of the proposed method is validated on the VAIS and RGB-NIR datasets and compared with the single-source image recognition method and other recognition methods in recent years under the identical experimental conditions.

3.4.1. Comparison with the recognition method of single-source images

Tables 2 and 3 list the recognition accuracy of different methods on the VAIS and RGB-NIR datasets. Three experiments were done to obtain the average and mean square error of recognition accuracy. It can be seen from Table 2, on the VAIS dataset, the recognition accuracy of the proposed method is greatly improved compared with that of the recognition method using only visible or infrared images, which is 0.034 higher than that of the method only using visible images and 0.082 higher than that of the method only using infrared images (IR). As observed from Table 3, on the RGB-NIR dataset, the recognition accuracy of the proposed method is 0.063 higher than that of the visible image recognition method, and 0.084 higher than that of the infrared image recognition method.

Tables 4 and 5 list the recognition results of different methods on the VAIS and RGB-NIR datasets. Here, the recognition accuracy of the proposed method is 0.936 and 0.818 on the VAIS and RGB-NIR datasets. It can be seen from Tables 4 and 5 that the proposed method had the highest average precision, average recall and average F1-score for the VAIS and RGB-NIR datasets. The reason is that as the effective classification features of the multimodal images are extracted through this proposed method, the complementary information of probability values is effectively utilized, and the recognition capability is enhanced after linear weighted decision fusion method.

Table 2. Recognition accuracy of different methods for the VAIS dataset.

Method	Accuracy		
	Visible	IR	Visible + IR
Improved CNN	0.902 ± 0.003	0.854 ± 0.002	–
Proposed method	–	–	0.936 ± 0.002

Table 3. Recognition accuracy of different methods for the RGB-NIR dataset.

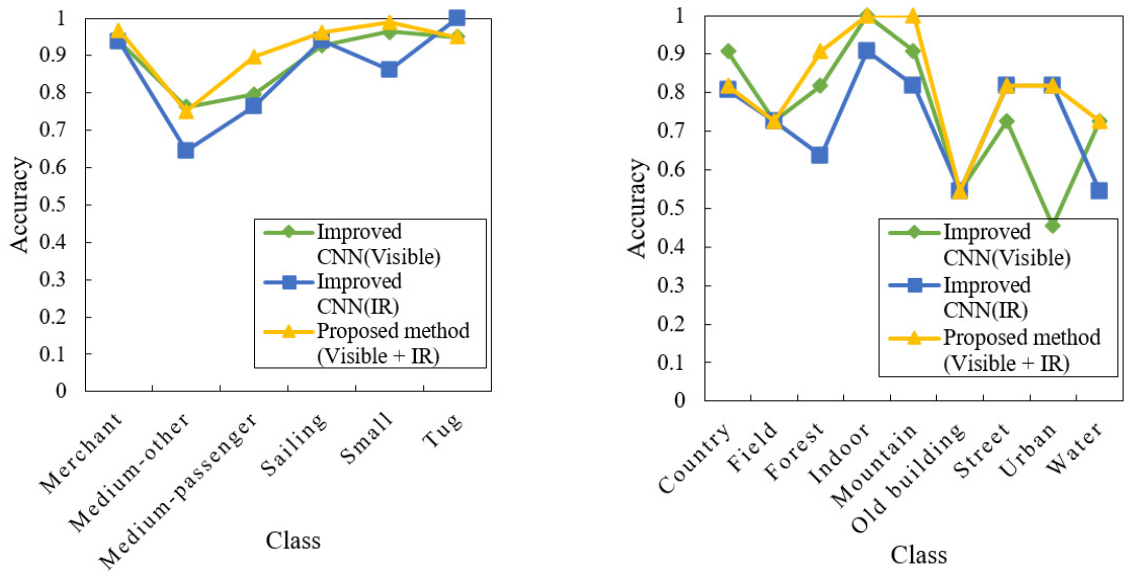
Method	Accuracy		
	Visible	IR	Visible + IR
Improved CNN	0.755 ± 0.006	0.734 ± 0.015	–
Proposed method	–	–	0.818 ± 0.01

Table 4. Recognition results of different methods for the VAIS dataset.

Method		Average Precision	Average Recall	Average F1-Score
Improved CNN	Visible	0.916	0.891	0.900
	IR	0.829	0.858	0.835
Proposed method	Visible + IR	0.955	0.920	0.935

Table 5. Recognition results of different methods for the RGB-NIR dataset.

Method		Average Precision	Average Recall	Average F1-Score
Improved CNN	Visible	0.768	0.758	0.753
	IR	0.777	0.737	0.739
Proposed method	Visible + IR	0.837	0.818	0.816



(a) VAIS dataset (b) RGB-NIR dataset

Figure 8. Recognition accuracy of different methods for each class.

3.4.2. Comparison of recognition performance with other recognition methods

Figure 8 shows the class recognition performance of different methods on the two datasets. As observed, on the VAIS dataset, the recognition performance of the proposed method is not improved for the tugboat, but it is improved for the other five categories. On the RGB-NIR dataset, the recognition performance of the proposed method is not improved for the country, but it has the best recognition accuracy for the other eight categories. The results show that the proposed method can improve the recognition performance well.

To further assess the advantages of the proposed method, it is compared with other recognition

methods. Figures 9 and 10 list the recognition accuracy of different methods for the VAIS dataset and RGB-NIR dataset, respectively. Figures 11 and 12 list the number of misrecognized multimodal images of different methods for the VAIS dataset and RGB-NIR dataset, respectively. Here, HOG + SVM method, LBP + SVM method, AlexNet, method [35] and method [10] only identify single source images. MOPDF represents maximum output probability decision fusion method, which takes maximum output probability as the criterion for decision processing. In method [20], multimodal image features extracted by CNN are concatenated fusion before recognition. In method [22], multilayer convolution features are first fused, and then the support vector machine posterior probability of each band is weighted fusion to achieve decision fusion recognition. SIFT [36] and AlexNet only identify single source images. Rgbi-SIFT [36] is a method that is used to extract SIFT features after infrared information is added to visible images. BoVW [36] is a recognition technology based on semantic feature extraction. Method [37] fused the visible and infrared images, then processed the fused images and obtained scene recognition results by combining the sparse recognition of the class dictionary. As observed, the recognition accuracy of different methods on the visible images is better than that on the infrared images because most infrared image resolutions are relatively low. The proposed method achieves the highest recognition accuracy and lowest number of misrecognized samples on the multimodal images compared with other recognition methods. This results show that the proposed method uses the dual CNN to extract more semantic features and detailed information of multimodal images and uses the linear weighted decision fusion method to process them, which can give appropriate weights to the visible and infrared subnetworks according to different multimodal input images. The proposed method effectively uses the respective advantages of multimodal images and improves the recognition accuracy.

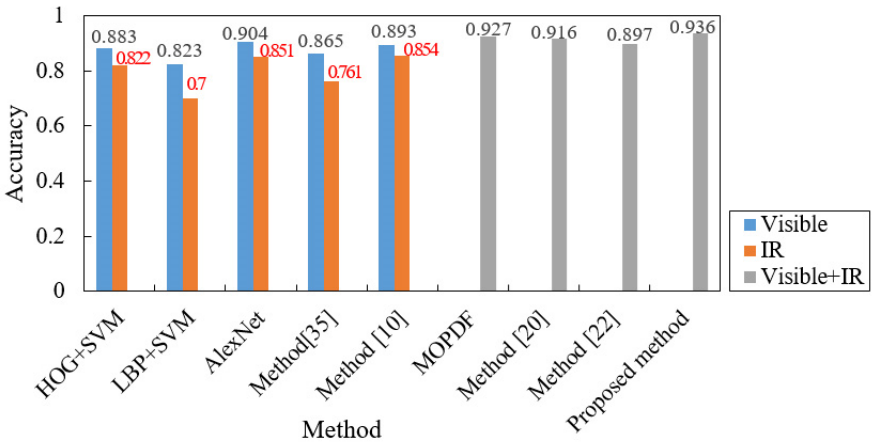


Figure 9. Comparison of recognition accuracy between proposed method and other methods on the VAIS dataset.

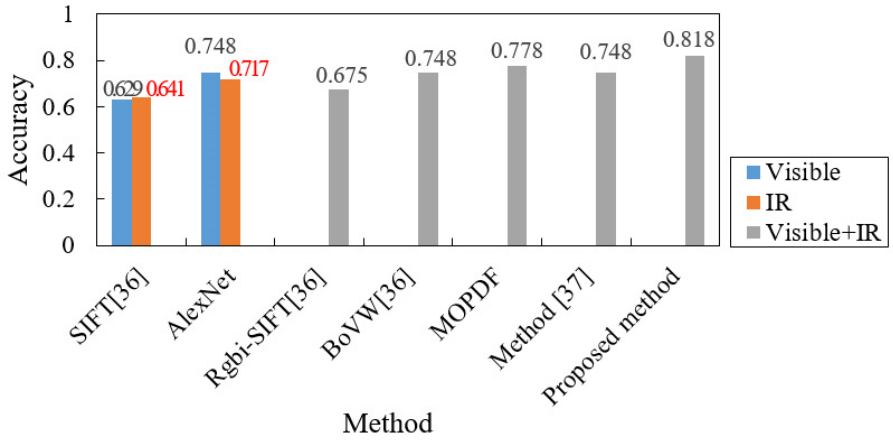


Figure 10. Comparison of recognition accuracy between proposed method and other methods on the RGB-NIR dataset.

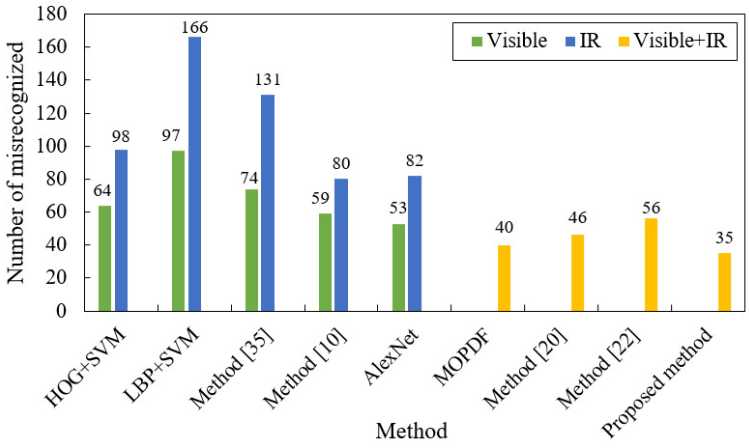


Figure 11. Comparison of the number of misrecognized multimodal images between proposed method and other methods on the VAIS dataset.

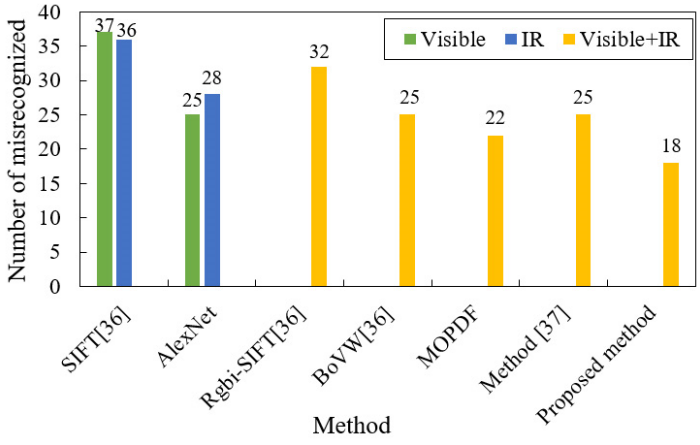


Figure 12. Comparison of the number of misrecognized multimodal images between proposed method and other methods on the RGB-NIR dataset.

In addition, to further verify the recognition performance of the proposed method, Tables 6 and 7 list the F1-scores of different methods for the VAIS and RGB-NIR datasets. As observed, the average F1-score of the proposed method is higher than that of other methods. For the VAIS dataset, MOPDF gives the highest F1-score for small ships, and AlexNet achieves the highest F1-score for merchant. While the proposed method attains the highest F1-score for all other four categories. For the RGB-NIR dataset, AlexNet achieves the highest F1-score for country and field, and MOPDF gives the highest F1-score for mountain. However, the proposed method attains the highest F1-score for the all other six categories. As the effective classification features of multimodal images are processed by linear weighted fusion method, the proposed method can effectively utilize the complementary information of multimodal images to more comprehensively represent the features, and further enhance the recognition ability.

Table 6. F1-scores of different methods for the VAIS dataset.

Method		Class						
		Medium-other	Merchant	Medium-passenger	Sailing	Small	Tug	Avg.Total
AlexNet	Visible	0.806	0.954	0.829	0.945	0.922	0.833	0.881
	Infrared	0.707	0.891	0.726	0.942	0.857	0.851	0.829
Method [35]	Visible	0.740	0.853	0.852	0.917	0.888	0.844	0.849
	Infrared	0.520	0.803	0.606	0.843	0.830	0.756	0.726
Method [10]	Visible	0.818	0.923	0.891	0.916	0.909	0.783	0.873
	Infrared	0.689	0.891	0.733	0.939	0.876	0.800	0.821
MOPDF	Visible + Infrared	0.814	0.947	0.875	0.978	0.936	0.974	0.921
Method [20]	Visible + Infrared	0.851	0.946	0.833	0.949	0.925	0.950	0.909
Proposed method	Visible + Infrared	0.832	0.953	0.938	0.978	0.935	0.974	0.935

Table 7. F1-scores of different methods for RGB-NIR dataset.

Method		Class									
		Country	Field	Forest	Indoor	Mountain	Old building	Street	Urban	Water	Avg.Total
AlexNet	Visible	0.833	0.842	0.857	0.778	0.800	0.667	0.800	0.467	0.800	0.760
	Infrared	0.643	0.842	0.842	0.609	0.917	0.632	0.609	0.720	0.667	0.720
MOPDF	Visible + Infrared	0.783	0.762	0.857	0.714	0.957	0.632	0.762	0.727	0.800	0.777
Proposed method	Visible + Infrared	0.783	0.800	0.909	0.786	0.917	0.667	0.818	0.818	0.842	0.816

Tables 8 and 9 list the feature extraction time per image of different methods for the VAIS dataset and the RGB-NIR dataset. As observed, the feature extraction time per image with the proposed method is slightly higher than that of the AlexNet and improved CNN due to the linear weighted decision fusion processing. For the VAIS dataset, the feature extraction time per image with the proposed method increased that of method [20] by 0.133 ms, but the recognition accuracy of the proposed method increased that of method [20] by 0.02. However, for VAIS dataset and RGB-NIR dataset, the feature extraction time per image with the proposed method reduced that of MOPDF by 0.017 and 0.02 ms, and

the recognition accuracy with the proposed method increased that of MOPDF by 0.009 and 0.04, respectively. The experimental results also show that after processing the effective classification features of multimodal images, better recognition performance is obtained. The feature extraction time per image of 0.333 and 0.192 ms is also relatively fast in practical applications. It is within the acceptable range.

Table 8. Feature extraction time per image of different methods for the VAIS dataset.

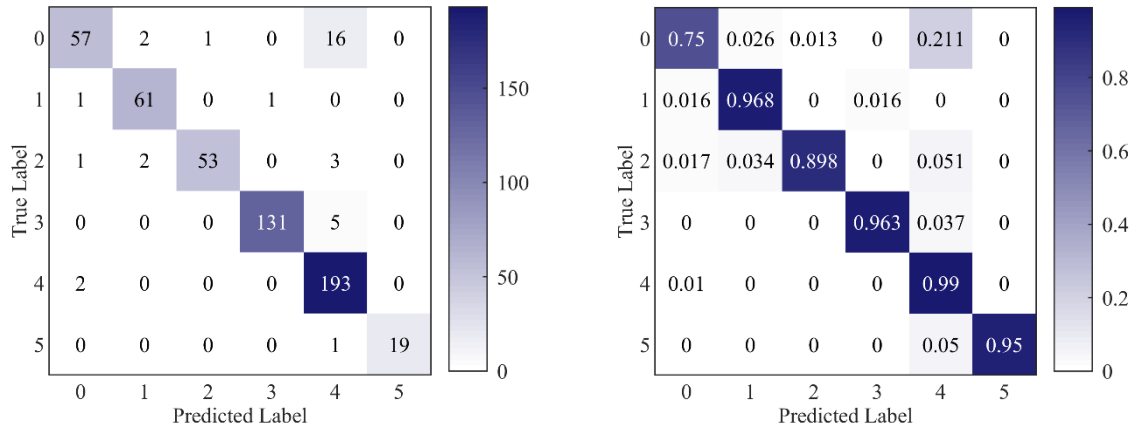
Method		Feature Extraction Time(ms)
HOG + SVM	Visible	8.978
	Infrared	8.854
LBP + SVM	Visible	23.395
	Infrared	22.582
AlexNet	Visible	0.104
	Infrared	0.062
Method [35]	Visible	0.189
	Infrared	0.608
Method [10]	Visible	0.053
	Infrared	0.053
Improved CNN	Visible	0.045
	Infrared	0.055
MOPDF	Visible + Infrared	0.350
Method [20]	Visible + Infrared	0.200
Proposed method	Visible + Infrared	0.333

Table 9. Feature extraction time per image of different methods for the RGB-NIR dataset.

Method		Feature Extraction Time(ms)
AlexNet	Visible	0.111
	Infrared	0.071
Improved CNN	Visible	0.081
	Infrared	0.061
MOPDF	Visible + Infrared	0.212
Proposed method	Visible + Infrared	0.192

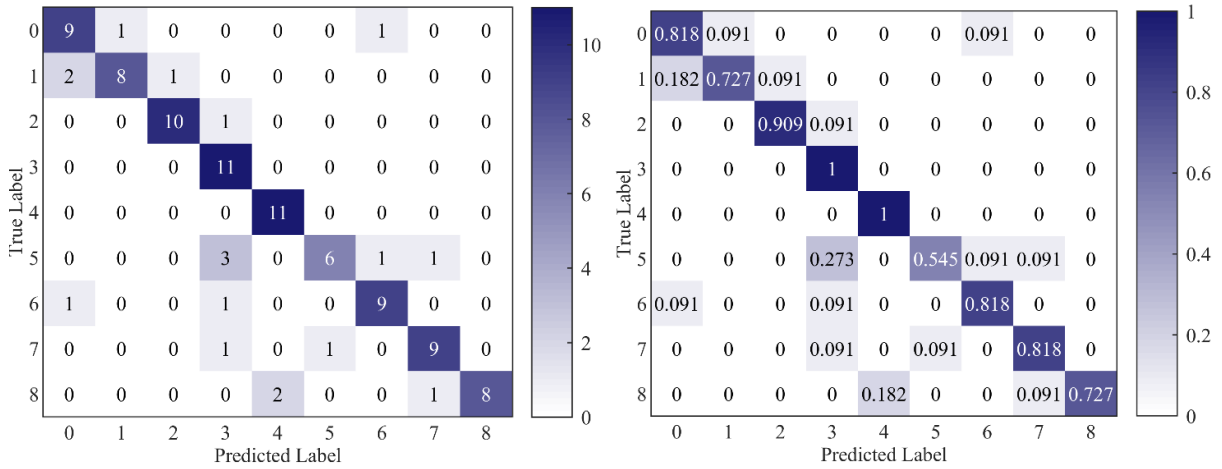
3.4.3. Recognition confusion matrix of the proposed method

Figures 13 and 14 depict the recognition confusion matrix of the proposed method for the VAIS and RGB-NIR datasets. In Figure 13, 0 is medium-other, 1 is merchant, 2 is medium-passenger, 3 is sailing, 4 is small, and 5 is tugboat. As observed, the key confusion occurred between classes 0 and 4 or between classes 2 and 4 or between classes 3 and 4. From the samples shown in Figure 6, we can observe that some medium-other ships and small ships show a noticeable resemblance, while some medium-passenger ships and sailing ships are blurry and also have similarities. In Figure 14, 0 is the country, 1 is the field, 2 is the forest, 3 is the indoor, 4 is the mountain, 5 is the old building, 6 is street, 7 is urban, and 8 is water. As observed, the confusion occurred between classes 3 and 5 because the interclass error is 0.273. As shown in Figure 7, some old buildings and indoor spaces exhibited a similarity.



(a) Confusion matrix (b) Confusion matrix normalization

Figure 13. Recognition confusion matrix of proposed method for VAIS dataset.



(a) Confusion matrix (b) Confusion matrix normalization

Figure 14. Recognition confusion matrix of proposed method for RGB-NIR dataset.

4. Conclusions

In this study, we presented a maritime ship recognition method for multimodal images based on CNN and linear weighted decision fusion. The proposed method first used a dual CNN method to learn the effective classification features of multimodal images. Then, the probability values classified by the softmax function were processed by linear weighted decision fusion and the recognition results were obtained. The dual CNN method could extract the effective classification features of the multimodal images, and the linear weighted decision fusion model could comprehensively consider the complementary information of the probability value of the multimodal images, thereby improving the ship recognition performance. Experimental results showed that, compared with the single-source image recognition and other recognition methods, the proposed method had the best recognition accuracy on the VAIS and RGB-NIR datasets, which were 0.936 and 0.818, respectively. In future

research, a multimodal image dataset with a larger sample can be examined to improve the maritime ship recognition ability of the proposed method.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the National Nature Science Foundation of China, grant number 51879211, the Hunan Provincial Education Department Science Research Youth Project of China, grant numbers 21B0800 and 22B0861, the Hunan Provincial Natural Science Foundation of China, grant number 2022JJ50148, the Hunan Provincial Education Department Science Research Key Project of China, grant number 22A0625, the Undergraduate Innovation and Entrepreneurship Training of Hunan province, grant number S202311528009 and the Guiding Planning Project of Hengyang, grant number 202222015678.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. L. Huang, F. X. Wang, Y. L. Zhang, Q. X. Xu, Fine-grained ship classification by combining CNN and Swin transformer, *Remote Sens.*, **14** (2022), 3087. https://doi.org/10.3390/rs14133087
2. T. Mustaqim, H. Tsaniya, F. A. Adhiyaksa, N. Suciati, Wavelet transformation and local binary pattern for data augmentation in deep learning-based face recognition, in *Proceedings of 10th International Conference on Information and Communication Technology*, (2022), 362–367. https://doi.org/ 10.1109/ICoICT55009.2022.9914875
3. Z. M. Zhuang, Z. J. Guo, Y. Yuang, Research on video target tracking technology based on improved SIFT algorithm, in *Proceedings of 7th International Conference on Electronics and Information Engineering*, (2016), 17–18. https://doi.org/10.1117/12.2265460
4. K. Sharma, P. K. Sarangi, L. Rani, G. Singh, A. K. Sahoo, B. P. Rath, Handwritten digit classification using HOG features and SVM classifier, in *Proceedings of 2nd International Conference on Advance Computing and Innovative Technologies in Engineering*, (2022), 2071–2074. https://doi.org/10.1109/ICACITE53722.2022.9823782
5. K. K. Tang, Y. X. Ma, D. R. B. Miao, S. Peng, Z. Q. Gu, Decision fusion networks for image classification, *IEEE Trans. Neural Netw. Learn. Syst.*, (2022), 1–14. https://doi.org/10.1109/TNNLS.2022.3196129
6. Z. Ma, G. D. Huang, Image recognition and analysis: A complex network-based approach, *IEEE Access*, **10** (2022), 109537–109543. https://doi.org/ 10.1109/ACCESS.2022.3213675

7. M. Xu, Z. Wang, X. M. Liu, L. H. Ma, A. Shehzad, An efficient pedestrian detection for realtime surveillance systems based on modified YOLOv3, *IEEE J. Radio Freq. Identif.*, **6** (2022), 972–976. https://doi.org/10.1109/JRFID.2022.3212907
8. T. W. Zhang, X. L. Zhang, J. Shi, S. J. Wei, A HOG feature fusion method to improve CNN-based SAR ship classification accuracy, in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, (2021), 11–16. https://doi.org/10.1109/IGARSS47720.2021.9553192
9. M. Z. Xu, Z. X. Yao, X. P. Kong, Y. C. Xu, Ships classification using deep neural network based on attention mechanism, in *Proceedings of 2021 IEEE/OES China Ocean Acoustics*, (2021), 1052–1055. https://doi.org/10.1109/COA50123.2021.9519897
10. Z. Z. Li, B. J. Zhao, L. B. Tang, Z. Li, F. Feng, Ship classification based on convolutional neural networks, *J. Eng.*, **21** (2019), 7343–7346. https://doi.org/10.1049/joe.2019.0422
11. J. W. Li, C. W. Qu, J. Q. Shao, Ship detection in SAR images based on an improved faster R-CNN, in *Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications*, (2017), 1–6. https://doi.org/10.1109/BIGSARDATA.2017.8124934
12. Y. Y. Wang, C. Wang, H. Zhang, C. Zhang, Q. Y. Fu, Combing single shot multibox detector with transfer learning for ship detection using Chinese Gaofen-3 images, in *Proceedings of 2017 Progress in Electromagnetics Research Symposium-fall*, (2017), 712–716. https://doi.org/10.1109/PIERS-FALL.2017.8293227
13. Y. Y. Wang, C. Wang, H. Zhang, Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images, *Remote Sens. Lett.*, **9** (2018), 780–788. https://doi.org/10.1080/2150704X.2018.1475770
14. M. Rostami, S. Kolouri, E. Eaton, K. Kim, Deep transfer learning for few-shot SAR image classification. *Remote Sens.*, **11** (2019), 1374. https://doi.org/10.3390/rs11111374
15. V. Ganesh, J. Kolluri, A. R. Maada, M. H. Ali, R. Thota, S. Nyalakonda, Real-time video processing for ship detection using transfer learning, in *Proceedings of* Third *International Conference on Image Processing and Capsule Networks*, (2022), 685–703. https://doi.org/10.1007/978-3-031-12413-6_54
16. Q. Q. Shi, W. Li, R. Tao, X. Sun, L. R. Gao, Ship classification based on multifeature ensemble with convolutional neural network, *Remote Sens.*, **11** (2019), 419. https://doi.org/10.3390/rs11040419
17. N. K. Mishra, A. Kumar, K. Choudhury, Deep convolutional neural network based ship images classification, *Def. Sci. J.*, **71** (2021), 200–208. https://doi.org/10.14429/dsj.71.16236
18. C. W. Wang, J. F. Pei, S. Y. Luo, W. B. Huo, Y. L. Huang, Y. Zhang, et al., SAR ship target recognition via multiscale feature attention and adaptive-weighed classifier, *IEEE Geosci. Remote Sens. Lett.*, **20** (2023), 4003905. https://doi.org/10.1109/LGRS.2023.3259971
19. F. Ucar, D. Korkmaz, A novel ship classification network with cascade deep features for line-of-sight sea data, *Mach. Vision Appl*., **32** (2021), 73. https://doi.org/10.1007/s00138-021-01198-2
20. K. Aziz, F. Bouchara, Multimodal deep learning for robust recognizing maritime imagery in the visible and infrared spectrums, in *Proceedings of the International Conference Image Analysis and Recognition 2018*, (2018), 235–244. https://doi.org/10.1007/978-3-319-93000-8_27
21. Y. Yang, K. F. Ding, Z. Chen, Ship classification based on convolutional neural networks, *Ships Offshore Struct.*, **17** (2022), 2715–2721. https://doi.org/10.1080/17445302.2021.2016271

22. X. H. Qiu, M. Li, G. M. Deng, L. T. Wang, Multi-layer convolutional features fusion for dual-band decision-level ship recognition, *Opt. Precis. Eng.*, **29** (2021), 183–190. https://doi.org/10.37188/OPE.20212901.0183
23. Y. H. Zhang, L. G. Li, Application of improved SqueezeNet in ship classification, *Transducer Microsyst. Technol.*, **41** (2022), 150–152+160. https://doi.org/10.13873/J.1000-9787(2022)01-0150-03
24. X. Du, J. Wang, Y. Li, B. Tang, Marine ship identification algorithm based on object detection and fine-grained recognition, in *Advanced Intelligent Technologies for Industry. Smart Innovation, Systems and Technologies*, (eds. K. Nakamatsu, R. Kountchev, S. Patnaik, J. M. Abe and A. Tyugashev), Academic Press, (2022), 207–215. https://doi.org/10.1007/978-981-16-9735-7_19
25. Z. L. Zhang, T. Zhang, Z. Y. Liu, P. J. Zhang, S. S. Tu, Y. J. Li, et al., Fine-grained ship image recognition based on BCNN with inception and AM-softmax, *Comput. Mater. Continua.*, **73** (2022), 1527–1539. https://doi.org/10.32604/cmc.2022.029297
26. L. Huang, F. Wang, Y. Zhang, Q. Xu, Fine-grained ship classification by combining CNN and swin transformer, *Remote Sens*., **14** (2022), 3087. https://doi.org/10.3390/rs14133087
27. W. L. Wang, X. D. Yang, B. Y. Zhang, J. S. Ma, P. Zeng, P. Han, Application of lightweight convolutional neural network in ship classification (in Chinese), *Laser Optoelectron. Prog.*, **60** (2023), 73–80. https://doi.org/10.3788/LOP213033
28. W. Sun, J. Yan, A CNN based localization and activity recognition algorithm using multi-receiver CSI measurements and decision fusion, in *Proceedings of the 2022 International Conference on Computer, Information and Telecommunication Systems*, (2022), 1–7. https://doi.org/10.1109/CITS55221.2022.9832983
29. W. N. Zhou, L. H. Sun, Z. J. Xu, A real-time detection method for multi-scale pedestrians in complex environment, *J. Electron. Inform. Technol.*, **43** (2021), 2063–2070. https://doi.org/10.11999/JEIT161032
30. J. L. Guo, Q. Liu, E. Q. Chen, A deep reinforcement learning method for multimodal data fusion in action recognition, *IEEE Signal Process. Lett.*, **29** (2022), 120–124. https://doi.org/10.1109/LSP.2021.3128379
31. M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, C. Kanan, VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums, in *Proceedings of the 2015 IEEE Computer Vision and Pattern Recognition Workshops*, (2015), 10–16. https://doi.org/10.1109/CVPRW.2015.7301291
32. M. Brown, S. Süsstrunk, Multi-spectral SIFT for scene category recognition, in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, (2011), 177–184. https://doi.org/10.1109/CVPR.2011.5995637
33. N. Saqib, K. F. Haque, V. P. Yanambaka, A. Abdelgawad, Convolutional-neural-network-based handwritten character recognition: an approach with massive multisource data, *Algorithms*, **15** (2022), 129. https://doi.org/10.3390/a15040129
34. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, (2012), 1097–1105. http://dx.doi.org/10.1145/3065386
35. K. Rainey, J. D. Reeder, A. G. Corelli, Convolution neural networks for ship type recognition, in *Proceedings of the SPIE 9844, Automatic Target Recognition XXVI*, (2016), 17–21. https://doi.org/10.1117/12.2229366

36. Q. S. Zhang, W. Li, L. Li, F. Zhang, H. T. Lang, Infrared and visible image fusion classification based on a codebookless model (in Chinese), *J. Beijing Univ. Chem. Technol*. (*Nat. Sci.*), **45** (2018), 71–76.
37. M. Wei, *HSV fusion of near-infrared image and visible image for scene recognition via sparse recognition using intra-class dictionary*, Master's thesis, Nanjing University of Posts and Telecommunications, 2019.

AIMS Press