Sentinel-2 Research on the Detection and Classification Methods of Maritime Ship Targets from Remote Sensing Images

Sentinel-2 Research on the Detection and Classification Methods of Maritime Ship Targets from Remote Sensing Images

Junjie He[1*], Yinan Lin[2], Fangzhe Shi[1], Jiajun Fu[1], Boning Chen[1]

[1]Guangzhou Maritime University, Guangzhou 510725, China
[2]Shenzhen University, Shenzhen 518060, China
Email: 476057168@qq.com

Abstract. There are problems such as low recognition accuracy and large classification error in the existing classification methods for ship identification based on optical remote sensing images. In this paper, we will analyze the characteristics of ships and determine the indicative factors for applying remote sensing to monitor ships in combination with optical remote sensing images. Using optical remote sensing image data, combined with U-Net and AttU-Net deep neural network models, we assist in extracting new remote sensing indices with strong generality and clear physical meaning, and establishing rules for monitoring ships, so as to establish a more general and clear physical meaning of the monitoring and identification method of remote sensing satellite images. The method is applied and evaluated with port optical remote sensing image data. The data show that compared with traditional machine learning methods, the accuracy of ship monitoring using U-Net and AttU-Net deep learning models in this paper reaches 89.04%, and the recall rate and accuracy rate are better than SVM. it shows that the model can detect ships effectively.

Keywords. Optical remote sensing images; Ship identification; Image classification recognition; Deep Learning; Deep Neural Network Model

1. Introduction

Marine natural resources are increasingly showing their strategic importance in today's national economic development, and are in a pivotal position in the major strategic decisions of China's maritime silk road. With the continuous development of modern measurement technology means, the monitoring tasks such as marine ship monitoring have also ushered in changes. Optical image-based ship detection technology started late and is still in the stage of theoretical and applied research on ship detection, which can be specifically separated from shore ship detection and near-shore ship detection [1]. Ya Yin [2] discussed the challenges of the current research and gave an outlook on the related development trends. Tengfei Wang[3] proposed a method combining superpixel representation and convolutional neural network methods for fast pixel-level detection of ship targets. Huili Wang [4] proposed a new edge-directional gradient histogram feature to characterize ship targets. Jie Huang [5] proposed a ship target detection method combining convolutional neural network (CNN) with support vector machine (SVM). Nan Wang [6] introduced a deep learning based target detection algorithm

algorithm into the field of ship detection.

2. Research Methodology

The existing remote sensing image pattern recognition technology has problems such as low recognition accuracy and large classification error in ship recognition and classification. Therefore, this paper proposes a ship recognition method based on the combination of deep learning and remote sensing technology, which has a large improvement in recognition and classification accuracy compared with the traditional supervised classification method, so that it can effectively achieve the detection of ships.

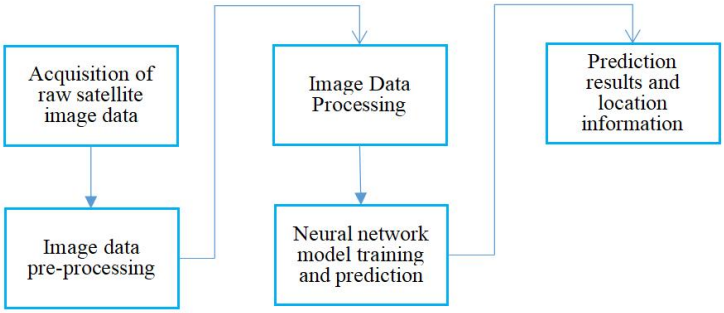


Figure 1. Overall flow chart

Figure 1 shows the method flow of this paper, first, obtain the original remote sensing satellite image data of the relevant area, then, pre-process the original data by ENVI professional geographic software for radiation correction, atmospheric correction, band fusion and other steps to get the area image that can be displayed normally, next, use visual interpretation combined with aerial orthophoto to select and crop the representative ship area and convert to Then, according to the ship's location information, we can use professional deep learning labeling tools to calibrate the ship area and get the image data that can be used for neural network training and validation, and after building a tuned neural network model for training, we can get a model that can predict the classification of the relevant ship area, and finally, we can match the model prediction results with the cropped images using professional geographic software to get the predicted area. Finally, the model prediction results are matched with the cropped images using professional geographic software to obtain the location information of the predicted area.

2.1. Pre-treatment Technology

The pre-processing step uses ENVI Professional Geography software to perform radiometric correction, atmospheric correction, band fusion and other operations on the raw data. ENVI Professional Geography software provides relevant general radiometric calibration tools and Landsat 8 remote sensing satellite atmospheric correction files, and ENVI Professional Geography software can complete the data pre-processing operations.

2.2. Deep Learning Models[7]

The models used in this paper include the U-Net [8] as well as the AttU-Net [9] deep neural network model. In the field of semantic recognition segmentation, the pioneering deep learning-based semantic segmentation algorithm is FCN (Fully Convolutional Networks for Semantic Segmentation) [10], and U-Net follows the principles of FCN with corresponding improvements to adapt it to simple segmentation tasks with small samples.U- Net network has gained a lot of attention because of its

advantages such as excellent recognition segmentation of fine objects in complex scenes and the relatively small amount of data required for training. In this paper, we adopt the U-Net model and its optimized version AttU-Net model for related research, where the AttU-Net model is based on the original U-Net model with the addition of an attention mechanism. The two neural network structures are shown in figure 2.

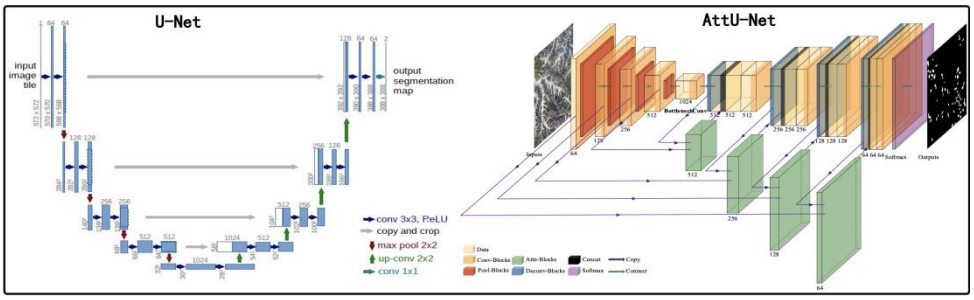


Figure 2. Network structure of U-Net and AttU-Net models

2.2.1 U-Net Model and AttU-Net Model. U-Net is a symmetric network structure, the structure is shown in the left part of figure 2, the left half is the downsampling step, the right half is the upsampling step, which corresponds to the encoder-decoder structure; and the four gray parallel lines in the middle are Skip Connection, its function is to fuse the feature information from the downsampling process in the upsampling process, and retain the features of the original data to a great extent, its fusion operation is relatively simple, only the channels of the feature map need to be superimposed, also known as Concat.

U-Net uses the pixel-level cross-entropy function as the Loss function, which is represented by Equation (1), where $l(x)$ is $\Omega \rightarrow \{1,\ldots,K\}$ is the true label of each pixel, $p_{l(x)}(x)$ is the softmax function of each pixel point, which is represented by Equation (2), where $a_k(x)$ denotes the first kth feature channel in the activation function value output by the pixel at $x \in \Omega$,Kdenotes the number of categories,$\omega(x)$is the boundary weight function, i.e., approaching a critical point pixel will have a higher weight, which is to enable the model to segment two targets that are very close to each other,$\omega(x)$is represented by Eq. (3), where$\omega_c(x)$is the weight feature of the $\Omega \rightarrow R$set mapping, while $d_1(x)$denotes the distance from pixel x to the nearest target boundary in the $\Omega \rightarrow R$et and $d_2(x)$denotes the distance from pixel x to the second nearest target boundary in the $\Omega \rightarrow R$set. In Eq. (3),ω_0 and$\sigma$are constants, and in the experiments in this paper, we set ω_0=10 as well as$\sigma \approx 5$。

$$E = \sum_{x \epsilon \Omega} \omega(x) log\ (p_{l(x)}(x)) \quad (1)$$

$$P_k(x) = \exp\ (a_k(x)) / \sum_{k'}^{K} \exp\ (a_{k'}(x)) \quad (2)$$

$$\omega(x) = \omega_c(x) + \omega_0 \cdot \exp\ (-\frac{(d_1(x)+d_2(x))^2}{2\sigma^2}) \quad (3)$$

2.2.2 AttU-Net Model. The AttU-Net model, on the other hand, combines the Attention Gates technique based on U-Net, which increases the sensitivity of the model to foreground pixels and suppresses the response of irrelevant background pixel regions, making the model better for segmentation of dense foreground targets.

The network structure of the AttU-Net model is shown in figure 2. Similar to U-Net, it is also a symmetric network structure, with the same down-sampling operation step as U-Net on the left half, and the innovative part compared to U-Net is the addition of Attention Gates to weight the most obvious part of the feature response in the up-sampling Concat step on the right half. Figure 3 shows the internal structure of the Attention Gates, which is based on the principle of fusing the features from the decoder stage with the corresponding features from the encoder stage through a 1×1×1size convolution, and then generating a weight mapα through a 1×1×1 size convolution combined with a Sigmoid activation function. whereαtends to obtain large values in the target organ region and smaller

values in the background region, which helps to improve the accuracy of image segmentation. In figure 3,$x^l$is the feature map from the encoder part, while $g$is the deep feature map from the decoder. The shallower texture shape and other feature maps and the deeper abstract feature maps are multiplied with the corresponding weight matrix $W$and then added, so that the model can namely retain the detail information and, at the same time, also retain the global semantic information to obtain a more complete Attention feature map, and finally the weight map$\alpha$is obtained after activation by the activation function, where $\alpha\epsilon[0,1]$, and finally the weight map$\alpha$is then multiplied with the feature map x^l of the encoder part.

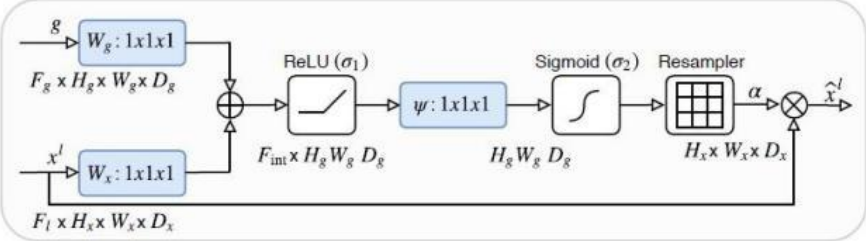


Figure 3. Internal structure of Attention Gates

3. Experimental Design and Analysis

3.1. Experimental Data and Study area

In this paper, representative ship classification training samples are selected by visual interpretation from the acquired remote sensing image data. The pre-processed images were cropped by ERDAS remote sensing satellite software, and a total of 116 images containing ship areas were cropped. The data set was divided into a training set and a test set by the ratio of 8:2, with 92 images in the training set and 24 images in the test set. In order to create the data set required for the U-Net network, this paper uses the labelme software to calibrate the existing 116 cropped images of the ship area and generate the corresponding mask data, the original images of the ship area and the corresponding mask are shown in figure 4.

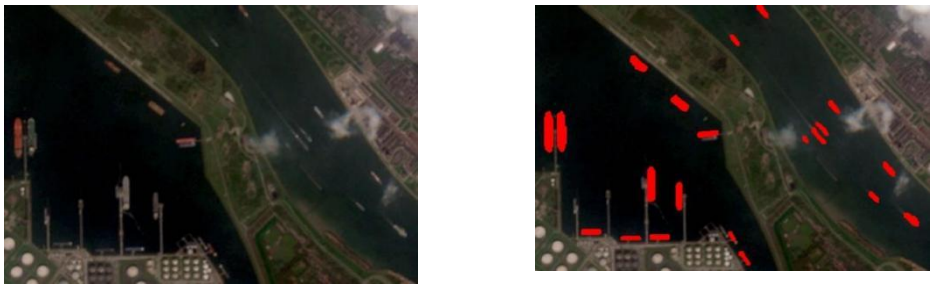


Figure 4. Experimental data of remote sensing satellite images and remote sensing satellite images annotated with labelme

3.2. Introduction of Validation Indicators

The validation metrics used in this paper include Accuracy, Recall, Precision, F1-score, Jaccard similarity coefficient and Dice similarity coefficient. (F1-score), Jaccard similarity coefficient (JS similarity coefficient), and Dice similarity coefficient (DS similarity coefficient), where Precision is the proportion of all predictions that are correctly classified, Accuracy is the percentage of correct predictions that are positive to all predictions, Recall is the percentage of correct predictions that are positive to all predictions that are positive, F1 is the balanced weighting of Accuracy and Recall, and the larger the F1 value, the better, Jaccard and Dice coefficients describe is the similarity between samples, which is often used in the field of medical segmentation, and the larger the value, the higher

the sample similarity.

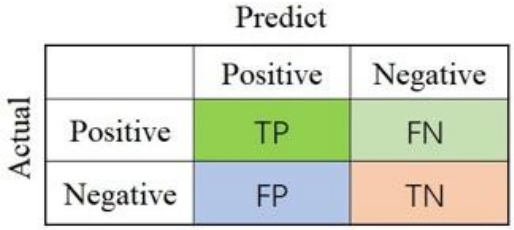


Figure 5. Map of evaluation indicators

The preparatory knowledge of the above index calculation formulas is represented by figure 5. True positive (TP) indicates that the prediction is positive and the actual case is positive; False positive (FP) indicates that the prediction is positive and the actual case is negative; True negative (TN) indicates that the prediction is negative and the actual case is positive; False negative (FN) indicates False negative(FN) means negative prediction and positive actual case, then the formulae of Accuracy, Recall, Precision, F1-score, Jaccard similarity coefficient and Dice similarity coefficient are (74)~(10) respectively, where The larger the value of Jaccard and Dice coefficients, the higher the sample similarity.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = \frac{2}{1/Precision+1/Recall} \quad (8)$$

$$JS(x,y) = \frac{|x \cap y|}{|x \cup y|} \quad (9)$$

$$DC(x,y) = \frac{2*|x \cap y|}{|x|+|y|} \quad (10)$$

3.3. Model Training

By building U-Net and AttU-Net deep learning network models for recognition, due to the small available training data set, this paper performs data enhancement operations on images before inputting them into the neural network. This paper uses conventional data enhancement techniques including randomly changing image hue, contrast, brightness, randomly cropping images, random horizontal flipping, random vertical flipping, etc. to expand the available data for model The amount of data available for model training is expanded to further improve the recognition accuracy and generalization ability of the model. The specific training parameters of the model are as follows: 150 training generations, learning rate of 0.002, learning rate decay starting at 70 generations and ending at 150 generations, linear decay, linear decay close to 0 from 70, and the model uses Adam's algorithm to update the network weights.

In the training phase, the model is trained and validated once to see if the model is overfitted or underfitted. As mentioned above, the entire data set consists of a training set and a test set, while the training set contains 92 images and the test set contains 24 images. In the training process, the training set and the test set are used for training and validation respectively, and the same data preprocessing is used for both training and test sets. The training and validation accuracy curves and loss curves are shown in figures 6 and 7. The comparison of figures 6 and 7 shows that the validation accuracy and loss curves of U-Net and AttU-Net models can be well fitted to the training progress and loss curves, and the models perform well without overfitting or underfitting. The model performance is stable after 150 generations of repeated training, as observed from the training curve graph.

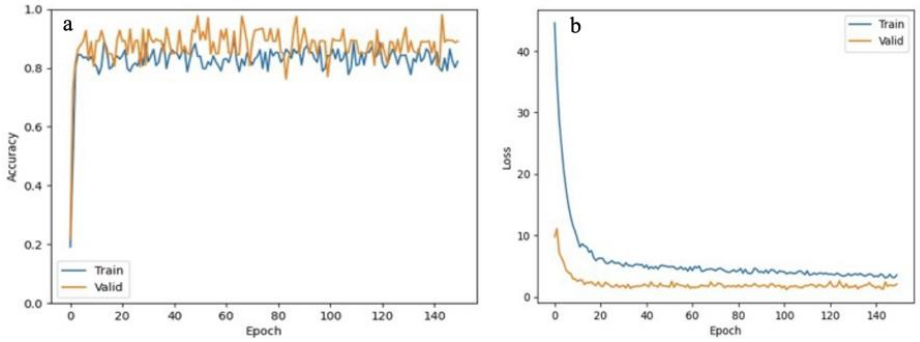


Figure 6. U-Net model training and validation performance graphs, a. Training and validation accuracy graphs; b. Training and validation loss graphs

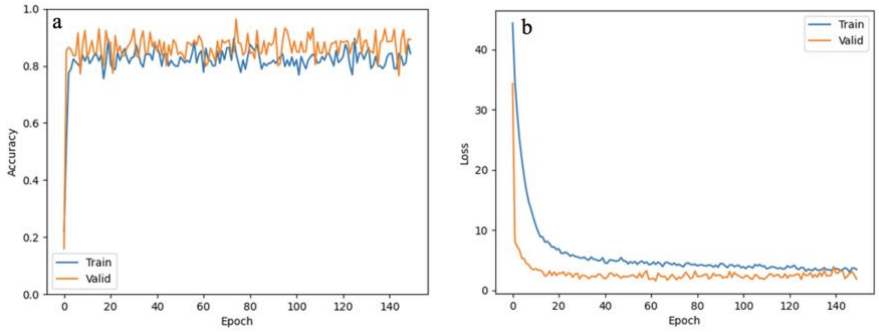


Figure 7. AttU-Net model training and validation performance graphs, a. Training and validation accuracy graphs, b. Training and validation loss graphs

After the model is trained, in the testing phase, the model is fed with the images to be tested, and its output corresponds to the predicted probability of each pixel, and the corresponding mask map is obtained by setting the segmentation threshold to 0.305, at which time the mask map is the predicted ship area map.

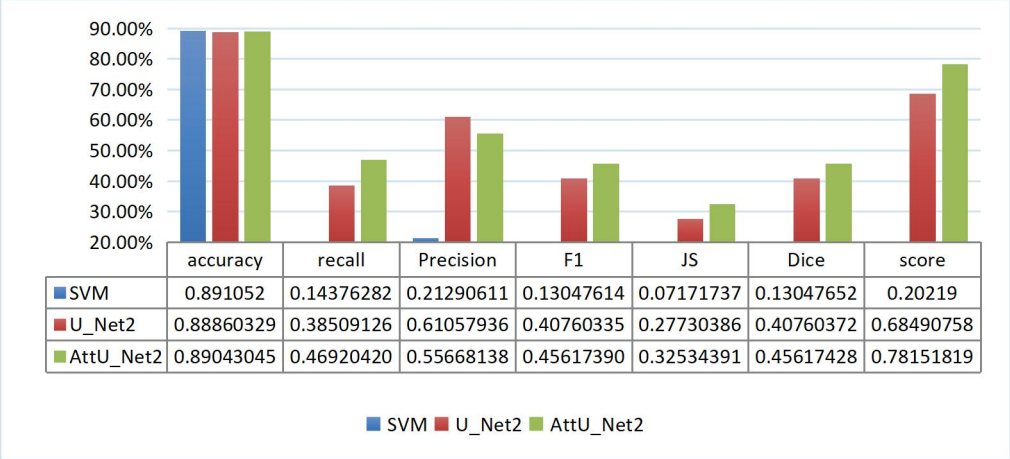
3.4. Analysis of Results

After the model training is completed, the parameter models with the most superior performance are saved separately, and the generalization ability and recognition accuracy of the model are tested using a test set of 24 remote sensing satellite images containing the ship region, and the images segmented from the test set are preprocessed using center cropping during the test. For comparison, SVM, which has a higher accuracy among the commonly used traditional supervised classification models, is selected in this paper for the test comparison, in which SVM selects 17 points of ship features. All the metrics in the table are more accurate the larger they are. The recognition effects of SVM, U-Net and AttU-Net models are shown in table 1, and the graphs of the three recognition results are shown in figure 8, where the red area is the ship area

Table 1. Model test data results

Methods (Train/Tes	SVM	U-Net (92/24)	Attention U-Net (92/24)
Accuracy	0.89105	0.88603	0.89043
Recall	0.14376	0.38509	0.46920
Precision	0.21291	0.61058	0.55668
F1-score	0.13047	0.40760	0.45174
JS coefficient	0.07171	0.27730	0.32534
DS coefficient	0.13047	0.40760	0.45617
Score (Jaccard + Di	0.20219	0.68490	0.78151

Table 2. Histogram of model performance



	accuracy	recall	Precision	F1	JS	Dice	score
SVM	0.891052	0.14376282	0.21290611	0.13047614	0.07171737	0.13047652	0.20219
U_Net2	0.88860329	0.38509126	0.61057936	0.40760335	0.27730386	0.40760372	0.68490758
AttU_Net2	0.89043045	0.46920420	0.55668138	0.45617390	0.32534391	0.45617428	0.78151819

From the test result data in table 1 and table 2, the SVM, U-Net, and AttU-Net models have 89.105%, 88.603%, and 89.043% accuracy, 14.376%, 38.509%, and 46.920% recall, 21.291%, 61.058%, and 55.668% accuracy, respectively, for the test set. F1 measurements were 0.1304, 0.4076, 0.4561, Jaccard coefficients were 0.0717, 0.2773, 0.4561, and Dice similarity coefficients were 0.1304, 0.4076, 0.4561, respectively.

Since the non-ship region occupies most of the pixels in the whole image, and SVM occupies most of the pixels for monitoring the non-ship region, it leads to a higher detection accuracy than the other two models, however, the correct rate of identifying the ship region cannot be simply expressed by Accuracy, the most important index is Recall as well as Precision, which can be found by the monitoring prediction of the three models in figure 8 U-Net, AttU-Net models monitor the ship region with a much higher correct rate than SVM, as observed in table 1 and table 2, it can be seen that, comprehensive various indicators, AttU-Net, U-Net and other deep neural network models have a much better recognition of the monitoring region than the traditional SVM classification algorithm, and AttU-Net incorporates the attentionattention mechanism. can better focus on the local ship area features, so its recognition performance is more superior.

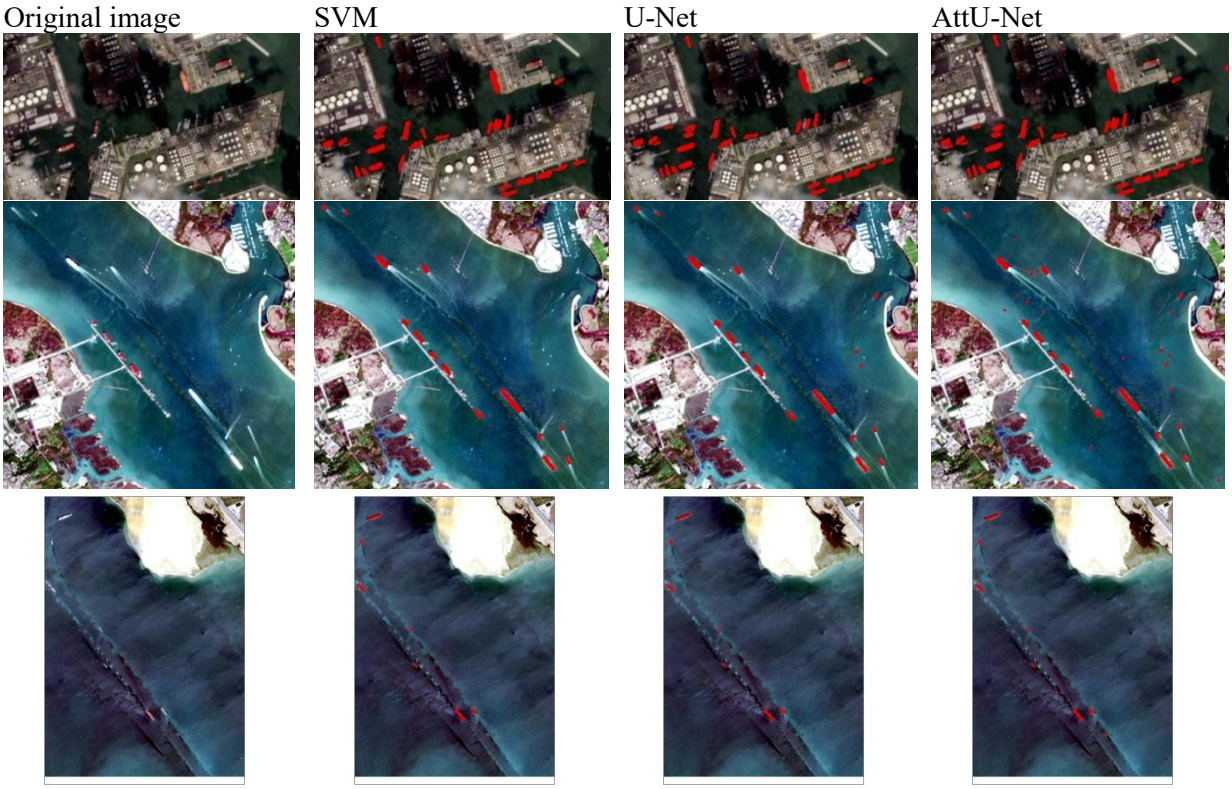


Figure 8. Monitoring effect of each model

4. Conclusion

Existing ship identification techniques are mainly based on manual identification, while emerging machine identification methods such as SVM and other algorithms have achieved good results, but their accuracy is not high enough and the recall rate is still low. Therefore, this paper proposes a fast monitoring and recognition technique based on the combination of deep neural network model and optical remote sensing images. Firstly, pre-processing techniques such as data cropping and enhancement are performed, followed by the construction of U-Net and AttU-Net recognition models to achieve effective ship identification and the accurate output of ship latitude and longitude coordinates through related techniques. It is shown by the data that the method can perform effective monitoring of ships. However, the existence of factors such as the amount of data available for training is too small, the characteristics of the data set are not obvious, and the number of image data periods is single, so there is room for further improvement of the recognition ability of the model.

Acknowledgments

This research content of this paper is supported by Guangdong Science and Technology Innovation Strategy Special Funds（pdjh2022b0392, pdjh2022b0394） and College Students' innovation and Entrepreneurship Projects(C2106001310, 202211106008).

Reference

[1]Zhang Zhixin. Remote sensing image ship detection and motion monitoring by geosynchronous orbiting satellite[D]. University of Chinese Academy of Sciences (Institute of Remote Sensing and Digital Earth Research, Chinese Academy of Sciences), 2017.

[2]YIN Ya,HUANG Hai,ZHANG Zhixiang. Research on ship target detection technology based on optical remote sensing images[J]. Computer Science,2019,46(03):82-87.
[3]Wang Tengfei. Research on deep learning ship detection technology for high-resolution remote sensing images[D]. Harbin Institute of Technology,2017.
[4]Wang Huili, Zhu Ming, Lin Chunbo, Chen Dianbing, Yang Hang. Ship detection in complex sea backgrounds in optical remote sensing images[J]. Optical Precision Engineering,2018,26(03):723-732.
[5]Huang Jie,Jiang Zhiguo,Zhang Haopeng,Yao Yuan. Remote sensing image ship target detection based on convolutional neural network[J]. Journal of Beijing University of Aeronautics and Astronautics,2017,43(09):1841-1848.
[6]Wang Nan. Deep learning-based ship detection and recognition [D]. Harbin Institute of Technology,2019.
[7] He Xiaohui, Qiu Fangbing, Cheng Xijie, Tian Wisdom, Zhou Guangsheng. Building target detection based on edge feature fusion for high resolution images[J/OL]. Computer Science:1-10.
[8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Springer, Cham, 2015.
[9] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning Where to Look for the Pancreas[J]. 2018.
[10] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.