Ship Classification by the Fusion of Panchromatic Image and Multi-spectral Image Based on Pseudo Siamese Lightweight Network

Mengyang Li1,2, Weiwei Sun1, Xuan Du3, Xiaohan Zhang1, Libo Yao1,*

1Naval Aviation University, Yantai, 264001, China
2Unit of the Chinese People's Liberation Army, Xuzhou, 050003, China
3Shanghai Academy of Spaceflight Technology, Shanghai, 201109, China

*502295312@qq.com

Abstract: The current rapid development of the remote sensing satellite industry provides a large amount of image data for ship classification tasks. Aiming at the problem of insufficient feature extraction of single source image, this paper designs a lightweight ship classification model based on the fusion of panchromatic image and multispectral image of pseudo Siamese network to extract image features more fully. First, establish a multi-source remote sensing image ship target classification dataset MPFS (MS and PAN Ship image Fusion Classification Dataset); secondly, send panchromatic images and multispectral images to the network through different convolutional layers, then design a multi-level feature extraction network for panchromatic images and an adaptive feature extraction network for spectral images respectively; finally, concatenate the features along the channel dimension and send them to the classification network.

Keywords: Ship classification; multi-source fusion; lightweight network

1. Introduction
With the continuous exploitation of marine resources, accurate classification of ship targets is of great significance for marine safety and traffic control. Remote sensing images have the advantages of wide coverage, high spatial resolution, and conformity with human visual representations, so become widely used in ship classification tasks.

In recent years, remote sensing satellites have developed rapidly, and both the quality and quantity...
of remote sensing images have been greatly improved. Now, there are many remote sensing satellites in the world, such as GF series, Resource series, QuickBird, IKONOS, etc. The satellites are equipped with different imaging equipment or imaging equipment of the same type but different resolutions to provide multi-source data for the same scene, which provides feasibility for data fusion. Most optical earth observation satellites and aerial photography systems provide both panchromatic images (PAN) and multispectral images (MS) [1], such as IKONOS with panchromatic image 1m and multispectral image 4m, GF2 with panchromatic image 1m and multispectral image 4m.

In the early days, [2] proposed a framework for fusion of remote sensing data based on Bayesian formulas to classify land objects; [3] proposed a general land cover classification model for multi-source remote sensing data based on Markov random fields; [4] introduced a fusion system, applying standard Kalman filter technology and fuzzy logic related technology for fusion of multi-source information to identify ship targets. However, the above feature extraction process needs to be completed manually, and are mostly based on certain conditional assumptions, which have some limitations.

In recent years, with the vigorous development of deep learning, its powerful feature extraction and representation ability has attracted researchers’ attention. The use of convolutional neural networks to extract features has been proven to achieve excellent results in remote sensing image classification tasks [5], and it has also been used in ship target classification tasks [6-7]. In the multi-source image fusion ship classification task, Chen [8] used optical images and SAR images to obtain information such as the length of the ship, the bow and the width of the ship, and then voted for recognition based on the detection results of each part. Liu [9-10] et al. carried out feature extraction on visible light, mid-wave infrared, long-wave infrared images of ships, then performed feature selection for training tests based on mutual information methods. However, the method of fusion of panchromatic image and multispectral image is only widely used in the task of land object classification [11-12], and there is no relevant research about ship classification.

The rigid body structure of the ship target is greatly affected by the weather, light and other photographing conditions, and with small inter-class differences as well as large intra-class differences, which adds the difficulty of classification task. How to effectively extract its features becomes a key point for classification. As is known, panchromatic images have high spatial resolution but only one spectral band; while multispectral images have multiple spectral bands but low spatial resolution. Accordingly, we can use panchromatic images to extract details such as contour edges and multispectral images to extract spectral information. The fusion of them can make full use of the complementary advantages of space and spectrum to get more effective target features.

Pseudo Siamese network [13] proves to be suitable for different but related inputs for classification tasks. Inspired by this, this paper proposes a lightweight network MPF (MS and PAN Fusion Network) for ship target classification based on the pseudo Siamese network fusion of panchromatic images and multi-spectral images. It can improve the classification accuracy by solving the problem of insufficient features provided by single source images; achieve the lightweight network compared with other networks; establish MPSF dataset by matching up panchromatic image and multispectral image obtained from the GF-2 remote sensing satellite. The proposed method achieved 75.59% accuracy rate on MPSC, which is improved to some degrees compared with other methods.

2. Proposed network

The overall architecture of the panchromatic and multi-spectral image ship target recognition lightweight network based on the pseudo-twin network designed in this paper is shown in Figure 1. It mainly includes the multi-level panchromatic image feature extraction branch and the adaptive multispectral image feature extraction branch. The features of the two features are connected in the channel dimension and then sent to the softmax classification network for training. The specific structure will be described in detail below.
2.1. **PAN densenet**

The panchromatic image has high spatial resolution, which means more detailed information of ship targets, such as the shape of the bow and the edge of ship contour. In convolutional neural network, the shallow layers can extract detailed information such as edge contours, and as the layers deepens, the extracted features also have higher abstraction. Inspired by the dense network [14], feature reuse through the connection of feature maps can enable the network to achieve better performance with fewer parameters and computational costs. In order to get more information of the panchromatic image, this paper proposes a multi-level feature extraction branch for panchromatic images, as shown in Figure 2.

![Architecture of PAN dense net](image)

**Fig. 2** Architecture of PAN dense net

First, after preprocess, images are sent to the first layer of convolutional network, and performed with maximum pooling operation. Then, send them to two compact modules with cascade between different feature layers, achieving feature reuse on channel dimension. The specific structure of the compact module will be described in detail. Finally, we perform the global maximum pooling on the output features with the aim of compressing the model and reducing the amount of calculation, and obtaining the most representative features of each channel.

A compact module is shown in the dashed box in Figure 2. This module includes a total of five levels of features $x_0, x_1, x_2, x_3, x_4$. $x_0$ represents the inputs from the former layer.

It is noted that, unlike dense networks, the compact module designed in this paper does not connect the output of the previous layer with each subsequent layer, but uses an interval connection method, with every two convolutional layer performs a feature connection.

Taking the first compact module as an example, the implementation process can be expressed as:

$$x_3 = H_3(x_0 \mid x_2)$$

$$x_4 = H_4(x_1 \mid x_3)$$

Among them, $H(\ )$ represents a nonlinear function, including operations such as BN, ReLU, pooling, and convolution, and $\mid$ means the feature map is concatenated along the channel dimension. In the convolution process, in order to keep the feature maps of different layers with the same size and realize the connection of the feature maps in different channel dimension, we set up the padding as 1,
and the kernel size as 3*3. The non-linear transformation process makes the network more portable and concise, accelerates the convergence speed, and increases the calculation speed.

2.2. MS adaptivenet

The multi-spectral image in this article has four band spectra. Compared with panchromatic image, it has higher spectral resolution, which means more sufficient spectral information. In this section, inspired by the classic network Vgg16[15], after many comparative experiments, we designed a spectral image feature extraction branch, as is shown in Figure 3. First, similar to the panchromatic image, the multi-spectral image is preprocessed and sent to the network. Secondly, design three groups of non-linear structures including convolution layer with 3*3 kernels, ReLU, and BN. We obtain 128, 256, and 512dimension feature maps respectively after three groups of non-linear structures. In order to retain the spectral information better and use the correlation of the original spectral information, the pooling layer is only added after the last convolution to speed up the convergence process. Finally, use the global maximum pooling to select the most representative feature on each feature channel.

![Architecture of MS adaptive net](image)

**Fig. 3 Architecture of MS adaptive net**

2.3. Training strategy

As there is no suitable pre-trained model available for the training dataset, it needs to be trained from the beginning, and the end-to-end training method is adopted. The parameter update is completed through back propagation.

The training loss is:

\[ L(Y, P) = -\log P(Y | P) = -\frac{1}{N} \sum_{i=0}^{N} \sum_{k=0}^{K-1} y_{i,k} \log p_{i,k} \]

Among them, \( Y \) represents the true probability, \( P \) represents the predicted probability, \( N \) represents the total number of samples, \( K \) represents the number of categories of the samples, here \( K=6 \),

\[ p_{i,k} = \frac{e^{w^{T}f_{j}+b}}{\sum_{j=1}^{6} e^{w^{T}f_{j}+b}}, \quad w \text{ is the classification network weight, } b \text{ is the bias, } f^{i} \text{ is the feature obtained by } f_{pAN} \text{ and } f_{sos} \text{ concatenate along the channel dimension.} \]

Since the fully connected layer can receive input feature vectors of fixed size, so the image size is uniformly mapped to 128*128 before training. Deep learning is a data-driven algorithm which must be supported by a large amount of data. In view of the high cost of labeling remote sensing images, this paper uses the data enhancement method of random horizontal flip, random vertical flip, and horizontal rotation of the image, which adds additional training data to make training more adequate. At the same time, the lightweight network structure designed in this paper is simpler than some classic networks, so it is prone to over-fitting problems, which leads to difficult training convergence and affects network performance. Therefore, in training step, we introduces the dropout layer after
acquiring the respective features of the two branches. According to parameter \( p \), the extracted features are randomly set to zero to increase the network robustness, reduce the amount of calculation, and improve the training efficiency.

3. Settings and experiments

The experimental platform is a computer equipped with ubuntu16.04 system and 1080Ti. It adopts the Pytorch deep learning framework. The parameter setting \( p \) is 0.5, the number of iterations is 250, the batch size is 8, the initial learning rate is 0.001, the learning rate is reduced to 20% every 50 epochs, and the ADAM training optimization method is chosen.

The MPSC dataset is derived from GF-2 satellite remote sensing images, which can simultaneously obtain panchromatic images and multispectral images of the same scene. After getting the slices of ship targets, we matchup the multi-source images. The resolution of the panchromatic image is 1m, while the multispectral is 4m which contains four band spectral of near-infrared, R, G, and B. MPSCDataset contains 6 types and 2632 groups of ship targets, which are Destroyers, Frigates, Combat boats, Bulk carriers, Container ships, Oil tankers. The dataset is divided into training set and testing set according to the ratio of 4:1, and the number of various samples is shown in Table 1.

| Class          | Destroyers | Frigates | Combat boats |
|----------------|------------|----------|--------------|
| Number         | 99         | 183      | 264          |
| Class          | Bulk carriers | Container ships | Oil tankers |
| Number         | 940        | 589      | 791          |

**Tab. 1** Numbers of different classes in MPSC dataset

In order to verify the effectiveness and superiority of the method proposed in this article, experiments will be conducted on MPSC dataset for the classification effect of multi-source image fusion and single-source image, of different network structures, and of different methods.

3.1. Comparison of multi-source image fusion and single-source image

In order to verify the effectiveness of the fusion method, we set up ablation experiments. The experimental methods and results are shown in the Table 2.

| Method       | Accuracy | Training time(s) | Testing time(ms) |
|--------------|----------|------------------|------------------|
| MSadaptive  | 64.88%   | 6.102            | 2.585            |
| PAN dense net| 71.77%   | 4.840            | 1.776            |
| MPF(propose d)| 75.59%  | 5.629            | 1.973            |

**Tab. 2** Comparison of multi-source image and single-source image

The above experiments show that the accuracy of the network structure constructed in this paper for panchromatic image and multi-spectral image fusion ship target recognition is much better than that of single-source image recognition. It reflects the complementarity of spatial and spectral information, and proves the effectiveness of the network. Considering the improvement in accuracy, the increase in training time and testing time is acceptable. At the same time, we can find that the classification accuracy of panchromatic images is higher than that of multispectral images, indicating that the amount of information provided by panchromatic images is more sufficient for the classification task.

3.2. Comparison of different network structures

In order to verify that the network structure designed for the two images in paper is superior, we tried different networks for different inputs. The experiments and results are shown in the Table 3.
Through experiments, it can be seen that training time and testing time are positively correlated with the complex structure of the network, but the recognition accuracy is not the same. The effect of the method proposed in this paper is generally better than that of Vgg16 and Resnet50. Panet (without dense block) means removing the interval connection in the compact module. Compared with this, PAN dense net has a very significant effect for extracting multi-level features of panchromatic images. Comparing MSnet (with pooling layer) and MS adaptive net, it can be seen that too many pooling layers in the network will cause the loss of spectral information. Thus we only need to add the pooling layer after the last convolution layer to preserve the spectral information to the greatest extent.

### 3.3. Comparison with other methods

Last, we compare our method with other methods, the results are shown in Table 4.

| Structure          | Accuracy | Training time(s) | Testing time(ms) |
|--------------------|----------|------------------|------------------|
| Siamese vgg16 shared | 67.80%   | 10.592           | 3.192            |
| Siamese vgg16      | 69.15%   | 11.519           | 5.916            |
| MRF[11]            | 68.32%   | 4.467            | 1.974            |
| Two-branchCNN[14]  | 69.90%   | 9.274            | 3.349            |
| MPF(proposed)      | 75.59%   | 5.629            | **1.903**        |

**Tab. 4 Comparison with other methods**

It can be seen that using Vgg16 as the baseline of the Siamese network to complete the fusion recognition task of this paper (parameter shared/no-shared), the accuracy is lower than the method proposed and the time cost is higher. At the same time, we can find that in the case of parameter shared, the recognition accuracy is lower, indicating that the types of information provided by the two
images are quite different, resulting in a low parameter sharing ratio. In particular, both MRF and Two-branch CNN are fusion networks for panchromatic images and multispectral images, compared with them, the proposed method has better performance on MPSC dataset.

4. Conclusions
This paper proposes a ship classification by the fusion of panchromatic image and multi-spectral image based on pseudo Siamese lightweight network. We design a multi-level feature extraction branch and a spectral adaptive extraction branch. The obtained features are concatenated just before the linear layer. This method is trained and tested on the self-built MPSC dataset, and the accuracy rate reaches 75.59%. Compared with other methods, the method proposed in this paper is superior in both accuracy and speed.

References
[1] Xiangchao M 2017 Variational fusion method of multi-source time-space-spectrum optical remote sensing image D. Wuhan University
[2] A H S Solberg, A K Jain and T Taxt July 1994 Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images IEEE Transactions on Geoscience and Remote Sensing 32 4 pp 768-778
[3] A H S Solberg, T Taxt and A K Jain Jan 1996 A Markov random field model for classification of multisource satellite imagery IEEE Transactions on Geoscience and Remote Sensing 341 pp 100-113
[4] Marc-Alain S, Eric L, and Christopher H M 2000 information fusion applied to ship identification for the recognized maritime picture Proc SPIE 4051 Sensor Fusion: Architectures, Algorithms and Applications IV
[5] N Liu, L Wan, Y Zhang, T Zhou, H Huo, and T Fang 2018 Exploiting convolutional neural networks with deeply local description for remote sensing image classification IEEE Access 6 pp 11215–11228
[6] Feng L, Tongsheng S, Xinxing M 2017 Feature fusion convolutional neural network multi-band ship target recognition J. Acta Optics 37(10) pp 248-256
[7] Shi Q, Li W, Tao R, Sun X and Gao L 2019 Ship Classification Based on Multifeature Ensemble with Convolutional Neural Network Remote Sens 11 p 419
[8] Wen C 2019 Research on typical target detection and recognition based on optical and SAR remote sensing image fusion D. Harbin Institute of Technology
[9] Feng L, Tongsheng S, Shaojun G and Jian Z 2017 Multi-band ship target recognition method based on feature level fusion J. Spectroscopy and Spectral Analysis 37(06) pp 1934-1940
[10] Feng L, Tongsheng S and Xinxing M 2017 Feature fusion convolutional neural network multi-band ship target recognition J. Acta Optics 37(10) pp 248-256
[11] R Gaetano, D Ienco, K Ose and R Cresson 2018 arXiv preprint arXiv:180611452 arxiv.org
[12] Xu X, Li W, Ran Q, Du Q, Gao L and Zhang B Feb 2018 Multisource Remote Sensing Data Classification Based on Convolutional Neural Network IEEE Transactions on Geoscience and Remote Sensing vol 562 pp 937-949
[13] Roopak S, Eduard S and James W 1993 Signature verification using a “siamese” time delay neural network J. International Journal of Pattern Recognition and Artificial Intelligence
[14] Gao H, Zhuang L, Geoff P, Laurens V D M and Kilian Q 2019 Weingerger IEEE Transactions
on Pattern Analysis and Machine Intelligence (IEEE TPAMI)

[15] Simonyan K and Zisserman 2014 A Very Deep Convolutional Networks for Large-Scale Image Recognition J. Computer ence

[16] He K, Zhang X, Ren S and et al 2015 J. Deep Residual Learning for Image Recognition