MULTI-SCALE SAR SHIP CLASSIFICATION WITH CONVOLUTIONAL NEURAL NETWORK

Xiaowo Xu, Xiaoling Zhang, Tianwen Zhang

University of Electronic Science and Technology of China

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553116

ABSTRACT

Ship classification in Synthetic Aperture Radar (SAR) images is significant but its application based on Convolutional Neural Network (CNN) has not been adequately studied. Considering that there will be the loss of SAR ship spatial information as the network deepening in CNN, which is a great obstacle for the further improvement of algorithm accuracy. Thus, to deal with the problem, in this paper, a novel multi-scale CNN (MS-CNN) is proposed. MS-CNN can utilize the multi-scale features to enhance the feature expression ability by the following three steps, namely flattening, integrating and classifying. As a result, the experiments on the OpenSARShip dataset show that MS-CNN can increase the classification accuracy by 4.81% than benchmark network.

***Index Terms*—**ship classification, Synthetic Aperture Radar (SAR), multi-scale Convolutional Neural Network (MS-CNN)

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is an active high-resolution microwave imaging radar with the advantages of all-weather and all-day working ability. SAR ship classification plays an essential role in SAR image intelligent interpretation. Especially, nowadays, more and more scholars pay attention to the ship classification in SAR images due to its huge application potential in marine traffic monitoring, ocean pollution monitoring, oil spill pollution detection, resources exploration and disaster risk assessment [1]-[5], etc.

In recent years, with the breakthrough success of deep learning technology in many fields [6], the methods of SAR ship classification using convolutional neural network (CNN) have been widely studied [7]-[10]. Carlos *et al*. [7] proposed a multiple input resolution CNN model (CNN-MR) to improve SAR ship classification performance. Lu *et al*. [8] utilized transfer learning to cope with the small training dataset problem of ship classification in high-resolution SAR images. Sao *et al*. [9] utilized a balanced sampling method to deal with the imbalanced phenomenon of the OpenSARShip dataset. Wu *et al*. [10] proposed a joint CNNs framework with an image super-resolution processing method to solve the problem of small SAR ship classification.

The above CNN-based methods achieved fairish SAR ship classification results on specific scenarios. However, these methods may face with great obstacles for the further improvement of algorithm accuracy, for the fact of that there will be the loss of SAR ship spatial information as the network deepening in CNN.

Thus, to address the above problem, in this paper, we proposed a novel multi-scale CNN (MS-CNN) to improve the SAR ship classification performance. Our idea is mainly inspired by the following facts.

To the best of our knowledge, although CNN-based multi-scale detectors have been extensively researched in the field of SAR ship detection, there are no similar CNN-based multi-scale classifiers in the field of SAR ship classification. What is more, it is a little unreasonable for CNN-based SAR ship classifiers to utilize only the features of the last convolutional layer in CNN, because there will be a loss for the spatial information of SAR ship as the network deepening in CNN, especially for small SAR ship. In fact, in the field of SAR ship classification, previous CNN-based studies only utilized the features of the last convolutional layer of CNN rather than multi-scale features.

As a result, in our MS-CNN, on the basis of the benchmark network, the multi-scale feature fusion is employed to enhance the feature expression ability by the following three steps, namely flattening, integrating and classifying. Finally, the experimental results confirm that MS-CNN can achieve a 4.81% improvement than benchmark network.

2. METHODOLOGY

2.1. Benchmark Network

First, we design a benchmark network, so that the effectiveness of the proposed method can be demonstrated by comparing the experimental results of benchmark network with the experimental results of MS-CNN.

Note that, benchmark network is composed of the backbone network of MS-CNN and a fully connected layer. The backbone network architecture of MS-CNN is shown in Table I. From Table I, it can be seen that each of our convolutional blocks ($L_1 - L_5$) is composed of a convolutional layer and a max-pooling layer. The input size or the output size is S×S×N, where S×S represents the size of feature map and N represents the number of convolution kernels. Besides, the size of convolution kernel is fixed at 3×3, as the network dee-

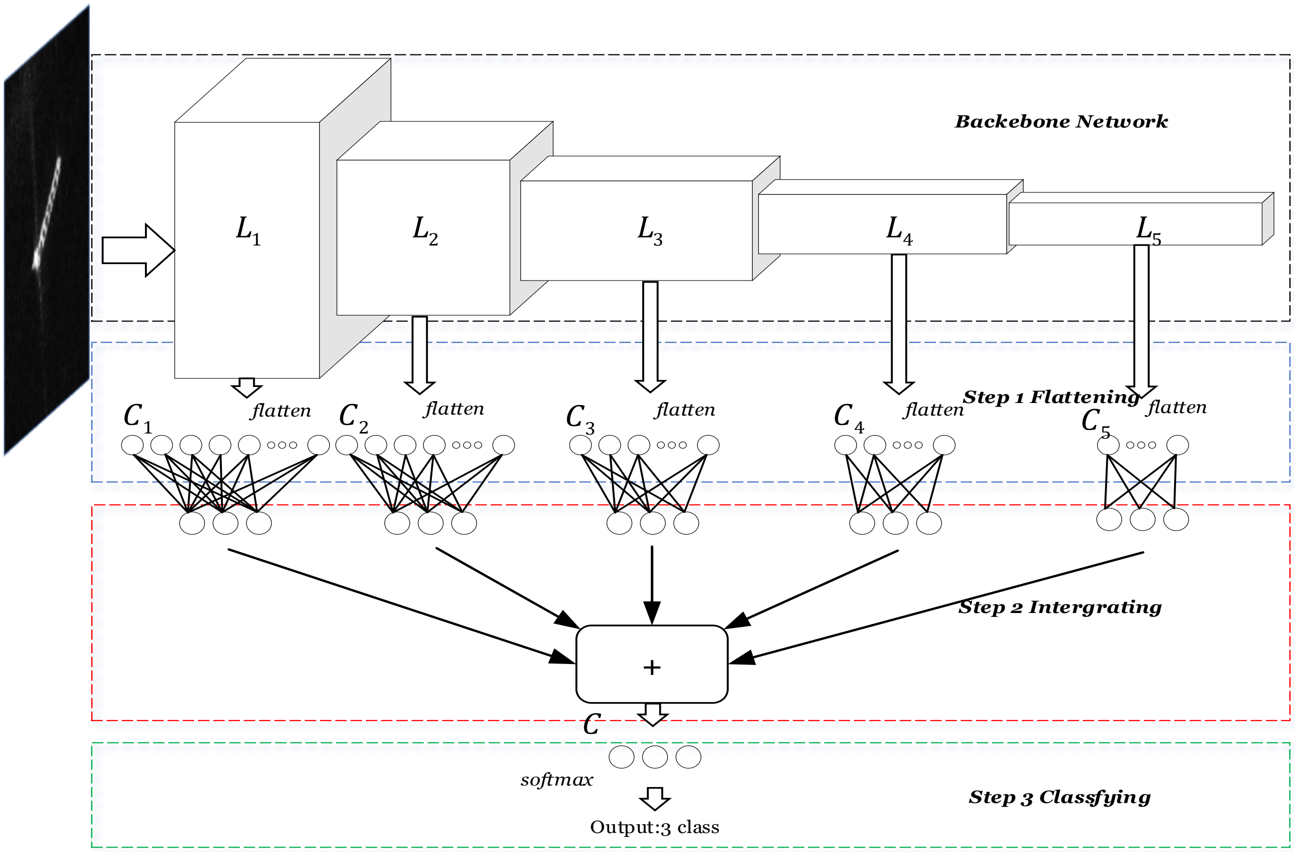

Fig. 1. Architecture of Multi-Scale CNN (MS-CNN).

TABLE I.
THE BACKBONE NETWORK ARCHITETURE OF MS-CNN

Block	Layer	Input Size	Output Size	Kernel@Stride
L_1	Conv.	128×128×3	128×128×8	3×3×8@1
	Pool.	128×128×8	64×64×8	@2
L_2	Conv.	64×64×8	64×64×16	3×3×16@1
	Pool.	64×64×16	32×32×16	@2
L_3	Conv.	32×32×16	32×32×32	3×3×32@1
	Pool.	32×32×32	16×16×32	@2
L_4	Conv.	16×16×32	16×16×64	3×3×64@1
	Pool.	16×16×64	8×8×64	@2
L_5	Conv.	8×8×64	8×8×128	3×3×128@1
	Pool.	8×8×128	4×4×128	@2

pening, the number of convolution kernels keeps increasing, i.e., from 8 to 128. The stride of convolution layers is set as 1 (marked by @1) while the stride of max-pooling layers is set as 2 (marked by @2).

Besides, the backbone network architecture of MS-CNN is also shown in the black dotted box of Fig. 1. Apparently, the backbone of the network consists of five convolutional blocks, which are respectively noted as $L_1 - L_5$ from the left to the right of the network. In addition, a softmax layer with a fully connected layer is directly adopted as the image classification layer of benchmark network.

Generally speaking, with the deepening of the network, semantic features extracted from the network are more abundant, but there will be more loss of the spatial features. For example, L_1 block is more suitable for spatial feature extraction, while L_5 block has a better semantic feature extraction ability. Besides, almost all proposed CNN-based SAR ship classification algorithms do not attempt multi-scale feature fusion.

Thus, on the basis of benchmark network, MS-CNN is proposed to solve the above problem.

2.2. Multi-scale CNN(MS-CNN)

Inspired by Feature Pyramid Network [11], in this paper, five different scales of networks are integrated to obtain the multi-scale features.

The architecture of MS-CNN is shown in Fig. 1. The main idea of MS-CNN is to utilize the multi-scale features to enhance the feature expression ability, which consists of three steps, namely flattening, integrating and classifying.

Next, the following details of three steps will be introduced.

2.2.1 Flattening

In order to obtain the multi-scale features, firstly we need to respectively flatten the output features of five blocks $\{L_1, L_2, L_3, L_4, L_5\}$ into 1-dimensional feature vectors, whose features experiments to call GPU for training acceleration.

TABLE II.
THE AMOUNT OF EACH CATEGORY IN THE OPENSARSHIP DATASET

Category	Training-Data	Testing-Data	All-Data
Bulk Carrier	338	328	666
Container Ship	338	808	1146
Tanker	338	146	484

(a) (b) (c)

Fig. 2. Three categories of SAR ship samples in the OpenSARShip dataset. (a) Bulk carrier; (b) Container ship; (c) Tanker.

are respectively named as $\{C_1, C_2, C_3, C_4, C_5\}$ (i.e., from features sizes of 64 × 64 × 8 to 32768 × 1 for the L_1 scale, from features sizes of 32 × 32 × 16 to 16384 × 1 for the L_2 scale, from features sizes of 16 × 16 × 32 to 8192 × 1 for the L_3 scale, from features sizes of 8 × 8 × 64 to 4096× 1 for the L_4 scale, from features sizes of 4 × 4 × 128 to 2048 × 1 for the L_5 scale).

2.2.2 Integrating

To start with, the features of each block processed by flattening will be inputted into five full connection layers respectively, so we can obtain 3×1-dimensional features vectors outputs of five different scales.

Secondly, we perform a simple addition operation on the five outputs from the last operation.

In short, the resulting fusion feature C can be defined by:

$$C = \sum_{L_{min}}^{L_{max}} FC_L\{C_L\} \tag{1}$$

where $L_{\min} = 1$ and $L_{\max} = 5$ respectively represent the index of convolutional blocks from left to right feature scale. FC_L is the L-th fully connected layer corresponding to L-th scale features.

2.2.3 Classifying

The obtained multi-scale fusion features are then inputted into a softmax layer to achieve the final classification prediction. In this step, since the inputs are the multi-scale fusion features, we will get a good classification performance.

3. EXPERIMENT

3.1. Experimental Platform

Our hardware platform is a Personal Computer (PC) with operating system of Linux 18.04, GPU model of NVIDIA RTX3090, CPU model of i7-10700K, memory size of 32G. We adopt Pytorch 1.7.0 [12] based on Python 3.6 language as the deep learning framework. We also use CUDA11.1 in our experiments to call GPU for training acceleration.

3.2. Dataset

The OpenSARShip dataset is widely used for SAR image intelligent interpretation [13]-[15]. Thus, we also use it to verify the effectiveness of the proposed MS-CNN. Taking account of class imbalance, we choose three major ship types: bulk Carrier, Container Ship and tanker, which keeps the same as Wang *et al*. [16]. Similarly, other dataset processing details are also the same as Wang *et al*. [16]. Table II shows the amount of each category in the OpenSARShip dataset. Fig. 2 shows three categories of SAR ship samples in OpenSARShip dataset.

3.3. Loss Function

In this paper, the cross entropy (CE) loss is used as the loss function to train MS-CNN in our experiments, as:

$$L_{cla} = -\frac{1}{N_{cla}} \sum_{i=1}^{N_{cla}} p_i \log(p_i^*) \tag{2}$$

where N_{cla} represents the number of training samples, i represents the index of the samples, p_{i} represents the i-th prediction, and p^*_{i} represents the corresponding ground truth label.

3.4. Training Details

We employ the Stochastic Gradient Descent (SGD) algorithm to train our network. The network input size is 128 × 128 and the batch 64 is adopted. Besides, we train network for 100 total epochs. We also set the learning rate as 0.0002, the weight decay as 0.0001 and the momentum as 0.9.

3.5. Evaluation Indices

The classification accuracy is defined by:

$$Accuracy = \left(N_{correct} / N_{all}\right) \times 100\% \tag{3}$$

where N_{correct} denotes the number of correctly classified samples, and N_{all} denotes the number of test set. Besides, we also utilize confusion matrix to intuitively show the classification results.

4. RESULTS

4.1. Classification Results

Table III shows the confusion matrix of the best accuracy in MS-CNN. Note that, we select the best 10 out of 100 epochs to calculate the mean and standard variance of accuracy. Apparently, from Table III, most targets can be classified correctly. However, we can see there is more misclassification between bulk carrier and container ship, because their texture features are more similar.

4.2. Comparison with Benchmark Network

Table IV shows the comparison between MS-CNN and benchmark network. Apparently, the proposed MS-CNN in this paper can make 4.81% accuracy improvement, because

TABLE III.
THE CONFUSION MARTIX OF MS-CNN IN THE OPENSARSHIP DATASET

Category	Bulk carrier	Container ship	Tanker	Accuracy (%)
Bulk carrier	218	70	40	66.46
Container ship	105	648	55	80.20
Tanker	14	10	122	83.56

TABLE IV.
COMPARISON BETWEEN MS-CNN AND BENCHMAEK NETWORK

Method	**Accuracy** (%)
Benchmark Network	72.39±0.75
MS-CNN(Ours)	**77.20±0.26**

MS-CNN can utilize the multi-scale features to enhance the feature expression ability. Thus, MS-CNN can achieve refined spatial information and semantic information, even when the texture features of different SAR ship types are similar. Finally, the SAR ship classification accuracy can achieve a great improvement.

5. CONCLUSION

In this paper, we proposed MS-CNN to address the problem of the spatial information loss of SAR ship as the network deepening in CNN. MS-CNN mainly consists of three steps, i.e., flattening, Integrating and classifying. Our experiments on the OpenSARShip dataset verify the effectiveness of MS-CNN. The experimental results show that the accuracy of MS-CNN can be improved by 4.81% compared with benchmark network, which indicates the advantage of MS-CNN in improving feature expression ability.

6. REFERENCES

[1] D. Velotto, M. Migliaccio, F. Nunziata and S. Lehner, "Oil-slick observation using single look complex TerraSAR-X dual-polarized data," *2010 IEEE International Geoscience and Remote Sensing Symposium*, Honolulu, HI, 2010, pp. 3684-3687.

[2] G. Suresh, G. Heygster, G. Bohrmann, C. Melsheimer and J. Körber, "An automatic detection system for natural oil seep origin estimation in SAR images," *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, Melbourne, VIC, 2013, pp. 3566-3569.

[3] S. Singha, D. Velotto, and S. Lehner, "Near real time monitoring of platform sourced pollution using TerraSAR-X over the north sea, "*Mar. Pollut. Bull*., vol. 86, no. 1, pp. 379-390, 2014.

[4] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 6, pp. 2218-2245, 2013.

[5] S. Voigt, T. Kemper, T. Riedlinger, R. Kiefl, K. Scholte and H. Mehl, "Satellite Image Analysis for Disaster and Crisis-Management Support," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1520-1528, June 2007.

[6] LeCun Yann, et. al, "Deep learning," *Nature*, vol. 521, pp. 436-444, May 2015.

[7] C. Bentes, D. Velotto and B. Tings, "Ship Classification in TerraSAR-X Images With Convolutional Neural Networks," in *IEEE Journal of Oceanic Engineering*, vol. 43, no. 1, pp. 258-266, Jan. 2018.

[8] C. Lu, and W. Li, "Ship Classification in High-Resolution SAR Images via Transfer Learning with Small Training Dataset," *Sensors*, vol. 19, no. 1, pp. 63, 2018.

[9] Q. Shao, C. Qu, J. Li, and S. Peng, "CNN based ship target recognition of imbalanced SAR image, " *Electron. Opt. Contr*., vol. 26, no. 9, pp. 90-97, 2019.

[10] Y. Wu et al., "Joint Convolutional Neural Network for Small-Scale Ship Classification in SAR Images," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019, pp. 2619-2622.

[11] Tsung-Yi Lin, P.D., Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie. "Feature Pyramid Networks for Object Detection." *arXiv*:1612.03144.

[12] Pytorch. Available online: https://pytorch-cn.readthedocs.io.

[13] J. Li, C. Qu, and S. Peng, "Ship classification for unbalanced SAR dataset based on convolutional neural network, " *J. Appl. Remote Sens*., vol. 12, no. 3, pp. 035010, 2018.

[14] Z. Huang, Z. Pan and B. Lei, "What, Where, and How to Transfer in SAR Target Recognition Based on Deep CNNs," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2324-2336, April 2020.

[15] Z. Huang, M. Datcu, Z. Pan, and B. Lei," Deep SAR-Net: Learning objects from signals," *ISPRS J. Photogramm. Remote Sens*., vol. 161, no. 2020, pp. 179-193, 2020.

[16] C. Wang et al., "Semisupervised Learning-Based SAR ATR via Self-Consistent Augmentation," in *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2020.3013968.