SAR MARITIME OBJECT RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORK

Yihang Zhi[1], *Bing Sun*[1], *Yi Xu*[1], *Jingwen Li*[1]

1 School of Electronic and Information Engineering, Beihang University, Beijing, China

ABSTRACT

Insufficient data of SAR target recognition task leads to low accuracy and poor generalization of model and the SAR imaging mechanism leads to the insignificant difference between ship targets, which make recognition difficult. To overcome the above problems, we propose a SAR maritime method using siamese networks for model pre-training. Siamese network produce sample pairs to ease training sample insufficiency, and output difference of sample pairs to help model learning heterogeneous difference. Then, transfer the pre-training parameters of feature exaction layer to an end-to-end model. Finally, the end-to-end convolutional neural network is obtained by fine-tuning the parameters with supervised information. Experimental results show that the SAR maritime target recognition method based on siamese network training can effectively improve the recognition accuracy under the training condition of a small number of samples.

Index Terms**—**SAR Maritime Object Recognition, Siamese Network

1. INTRODUCTION

With the development of synthetic aperture radar (SAR) imaging technology, SAR image is widely used in the intelligent identification of maritime targets, which plays an important role in military and civil fields, such as maritime traffic scheduling, illegal fishing and hunting monitoring, abnormal behavior detection[1]. In digital warfare, maritime target identification makes an indispensable contribution to battlefield analysis, precision guidance, military situation early warning and other military fields. Therefore, maritime target recognition is the focus of SAR image automatic target recognition (ATR). Nevertheless, the insufficient training data and speckle noise hinder SAR ATR performance.

Fortunately, the rise of deep learning drives the improvement of SAR image recognition performance. With the advantages of big data, GPU development and transferable model, frameworks represented by convolutional neural networks is effective in SAR image recognition.[2] However, SAR ATR tasks have far fewer labels and images than optical tasks, resulting in poor performance of deep learning model. Therefore, we use siamese network to produce sample pair in training, help model to learn more information via computing distance of training pair.

As imaging mechanism of SAR image is different from that of optical image, there are a lot of speckle noise in SAR image and the resolution of SAR image is much smaller than that of optical image, so the SAR image of the same target has little difference. Researchers[4] have tried to solve these problems, most of them would like to shorten the distance within the same class and increase the distance between different classes. In this paper, a training method based on siamese network is proposed, which takes feature distance of different classes as learnable errors. Siamese network[6] was first proposed by Bromley, aiming to solve the problem of image matching. Siamese network can find out the difference of similar targets[7]. Experiment results show that this training method can help deep learning model to achieve excellent recognition accuracy in SAR ATR tasks.

This paper is organized as follows: we introduce the pre-training method in Chapter 2, use siamese network training method to pre-train the convolutional neural network, then transfer the trained parameters to the end-to-end convolution neural network model to handle SAR maritime recognition task. The performance of this method on the GaoFen-3 satellite dataset is summarized and analyzed in Chapter 3. Finally the paper is summarized in Chapter 4.

2. METHODOLOGY

Our method is composed of pre-training stage and fine-tuning stage as shown in Figure 1. In the pre-training stage, samples are paired randomly. Then features of sample pair extracted by feature extraction layer and update parameters with cross-entropy loss. In the fine-tuning stage, transfer the parameters of feature extraction layer to new convolutional neural network and frozen them. The rest of layers are trained with labeled sample.

2.1 Pre-Training Stage

In the target recognition task, the traditional convolutional neural network training method only inputs one image at a time, and the learned knowledge is limited to

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883367

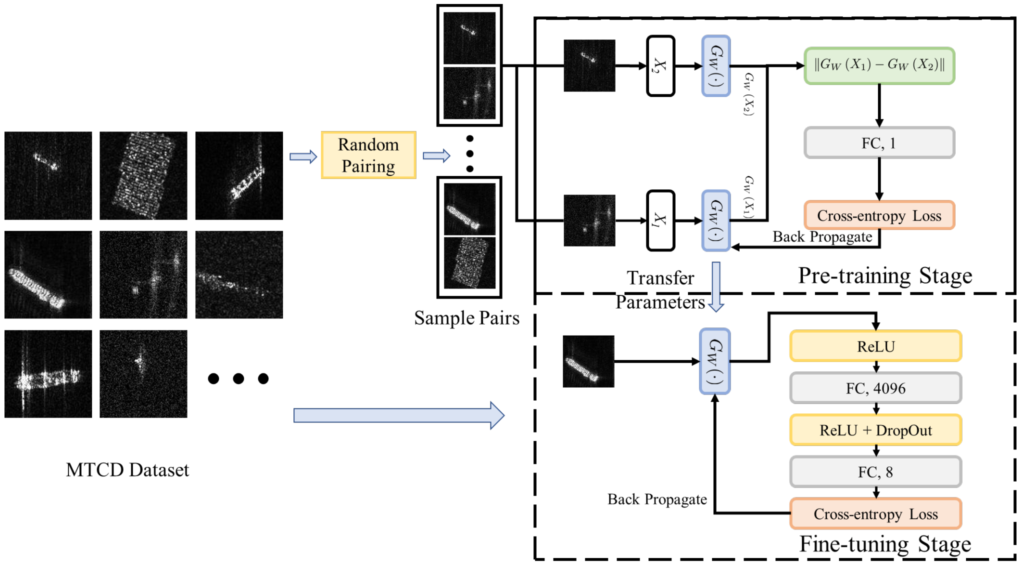


FIGURE 1. Pipeline of our approach.

one input image. The advantage of pre-training with siamese network is that it accepts two image inputs at a time, and updates the network parameters by learning the sample feature distance. The model can obtain more knowledge from different input sample pairs, so siamese network training allows model to learn more knowledge, especially the inter-class difference knowledge, furthermore, alleviate the problem of excessive data demand for deep learning methods.

Assuming a dataset contains N class, and each class contains M samples, theoretically, pair any two different samples in the dataset, the total number of sample pairs can be compute as follow:

$$N_{pairs} = C_{M \cdot N}^{2} = \frac{M \cdot N(M \cdot N - 1)}{2} \tag{2.1}$$

These sample pairs are composed of homogeneous sample pairs with label 1 and heterogeneous sample pairs with label 0, where the number of homogeneous sample pairs N_{same} is:

$$N_{same} = N \cdot C_{M}^{2} = \frac{M \cdot N(M - 1)}{2} \tag{2.2}$$

The number of heterogeneous sample pairs N_{diff} is:

$$N_{diff} = M^{2} \cdot C_{N}^{2} = \frac{M^{2} \cdot N(N - 1)}{2} \tag{2.3}$$

Therefore, we can be expanded the original $M \cdot N$ samples to $\frac{M \cdot N(M \cdot N - 1)}{2}$ sample pairs, and the theoretical expansion factor N_{avg} is:

$$N_{aug} = \frac{N_{pairs}}{M \cdot N} = \frac{\frac{M \cdot N(M \cdot N - 1)}{2}}{M \cdot N} = \frac{MN - 1}{2} \tag{2.4}$$

So training with siamese network can make full use of sample pairs to learn more knowledge.

The siamese network structure is shown in Figure 2. We used Convolutional neural network (ConvNet) as feature extraction layer $G_W($, with 4 convolution layers, and 1 full connected layers, because it has few parameter and easy to deploy in SAR system. The vector distance of feature is represented by L2 norm as (2.6). Finally, we obtain the trained $G_W($ as feature extraction layer, which will be use in end-to-end SAR maritime ATR model.

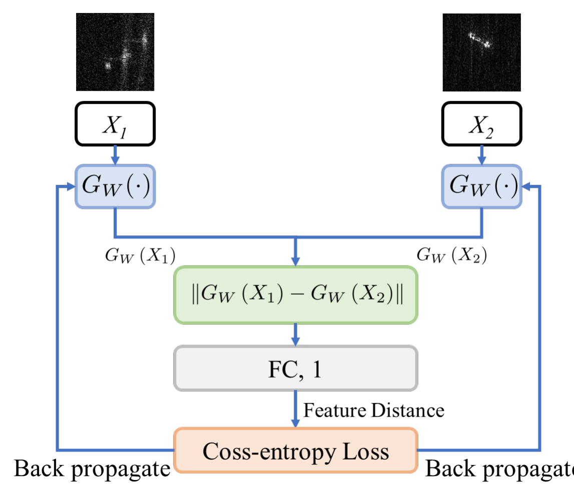


Figure 2. Siamese network structure using ConvNet as feature extraction layer.

The corresponding feature vectors are obtained after the sample pairs pass through the weight sharing of two ConvNets, and the L2 norm of feature vectors is computed as (2.5). Finally, the loss is compute by cross-entropy loss.

$$d = \sqrt{(G_W(X_1) - G_W(X_2))(G_W(X_1) - G_W(X_2))^T} \quad (2.5)$$

SAR ATR task affected by speckled noise, and does not perform well in CNN. Traditional convolutional neural network compute the training loss with output class and label, it learn knowledge from the homogeneous sample, but the differences between different species are not fully utilized to widen the distance between different class. Since the loss function of siamese network is related to the difference between the two sample features, in the process of training convergence, the feature distance extracted from images of the same class will decrease. On the contrary, the feature distance of different class will increase, showing a clustering trend. Therefore, it can overcome the problem that SAR images are too similar.

2.2 Fine-Tuning Stage

As the trained siamese network can not output sample class but the result of sample match, it cannot handle end-to-end SAR ATR tasks. Therefore, we transfer the trained parameters of feature extraction layer to an untrained ConvNet. Finally, we obtain an end-to-end convolutional network with pre-training parameters, as shown in Figure 2. The 5 convolutional layers and the first full connected layer have pre-training parameters.

Parameters of the last two full connected layers are random initialized. Therefore, supervised fine-tuning of the last two full connected layers is required. In fine-tuning phase, pre-training paraments are frozen, only the last two full connected layer are trained. In this way, the parameters learned by siamese network are retained, and the fine-tuning model can handle an end-to-end task.

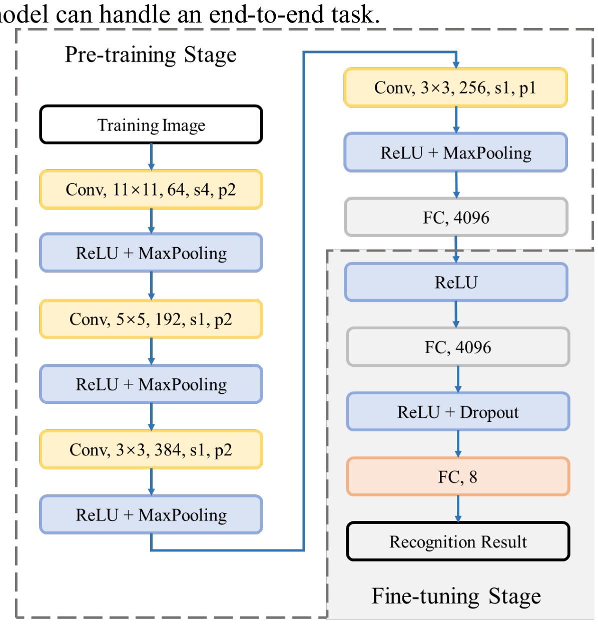


Figure 2. End-to-end Convolutional Neural Network consisting of pre-training part and fine-tuning part.

3. EXPERIMENTS

3.1. Dataset

In experiments, analysis is executed on MTCD dataset. Target distribution of Maritime Target Classification Dataset (MTCD) of Gaofen-3[11] is shown in Table 1, with 8 types of ships and 3225 images.

Table 1. Category Distribution of MTCD

Description	Training Set	Test Set
Boat	390	104
Cage	295	94
Cargo	400	154
Container	200	54
Tower	290	72
Platform	280	55
Tanker	312	76
Windmill	355	94
Total	2522	688

MTCD is divided into training data set and test set in a ratio of 9:1. Then 100%, 80%, 60%, 40% and 20% data used for training are extracted and divided into different training data, we call them MTCD1, MTCD0.8, etc. Each training data set was divided into training set and test set in a ratio of 7:2.

As Siamese network would produce sample pair, we ensure the sample pairs of datasets is similar by setting the extend times of different dataset according to (2.4). Extend results are shown in Table 2, and the number of positive and negative samples is kept equal in order to prevent sample imbalance.

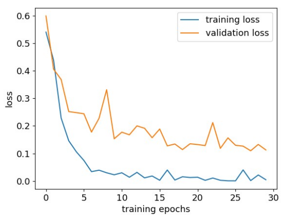
Table 2. Extend Results of MTCD Dataset

Training Set	MTCD 0.2	MTCD 0.4	MTCD 0.6	MTCD 0.8	MTCD 1
Extend Times	14	8	6	4	4
Sample Paris	22120	25280	28434	25276	31596

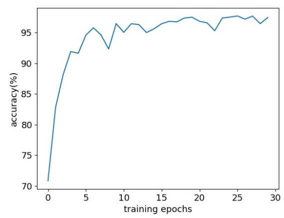
3.2. Experimental Results

We trained siamese network for 13 epochs, with batch size of 64. With the use of Adam optimization, the weight decay was set to 1e-5. The initial learning rate was 1e-4. Then we fine-tuned the end-to-end ConvNet for 60 epochs. Experiments are conducted with NVIDIA GTX 1080 where the Pytorch framework is implemented to perform our experiments.

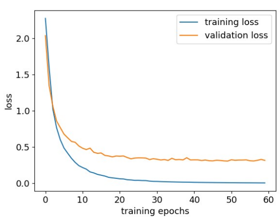
Taking MTCD-0.4 as an example, the training results of our method are shown in Figure 4. In the pre-training stage, the siamese network achieved convergence quickly in 30 epochs, and the accuracy is close to 90%, which shows that the pre-training of siamese network is worked.



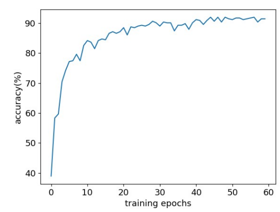
(a) Pre-training loss



(b) Pre-training accuracy



(c) Fine-tuning loss



(d) Fine-tuning accuracy

Figure 4. Training Results with MTCD-04 dataset include pre-training phase and fine-tuning phase. (a), (b) is pre-training loss and sample match accuracy. (c), (d) is fine-tuning loss and SAR maritime target recognition accuracy.

The comparison of different training methods and different training sample are shown in Table 3. The proposed method achieves high accuracy on MTCD. Compared with the deep learning training method, the smaller the amount of data, the more obvious the improvement of the proposed method.

Table. 3 Comparison of Our method and deep learning training method on MTCD

Training Set	MTCD-0.2	MTCD 0.4	MTCD-0.6	MTCD-0.8	MTCD-1
ConvNet	81.0%	89.1%	92.0%	95.7%	97.0%
Ours	**85.0%**	**89.9%**	**92.3%**	**95.8%**	96.7%

Result of MTCD-0.2 with least amount of data, the accuracy of our method is improved by 4%. However, in the case of large data volume, the effect of the proposed method is not significantly different from that of the end-to-end ConvNet method.

4. CONCLUSION

We proposes a training method based on siamese network for maritime target recognition for SAR images with small samples to overcome sample insufficient problem. The model parament is pre-trained with siamese network and sensitive to class information. After the siamese network converges, the feature extraction layer is saved and attach full connected layers at the end to obtain the end-to-end convolutional neural network for SAR ATR task. In the experimental stage, Gaofen-3 Maritime target classification dataset were used to conduct experiments under different amount of sample. Compared with the traditional training method, we verify the effectiveness of our method. Experimental results show that the accuracy of the proposed method is improved by 4.0% on MTCD dataset, and the accuracy of the proposed method is superior to the classic network under the condition of small samples. However, as the number of sample increases, the error transmission of the step training method leads to a decline in accuracy. If the dataset is small, the training strategy in this paper can be used to solve the problems of insufficient training samples and.

5. REFERENCES

[1] H. Fu, G. Song, and Y. Wang. "Improved YOLOv4 Marine Target Detection Combined with CBAM." *Symmetry*, vol 13, no. 4, pp. 623, 2021.

[2] Y. Duan., F. Liu, and L. Zhang. "SAR Image segmentation based on convolutional-wavelet neural network and markov random field." *Pattern Recognition,* vol. 64, pp. 255-267, 2016.

[3] Q. Fan, F. Chen, M. Cheng, S. Lou, R. Xiao, B. Zhang, C. Wang, and J. Li. "Ship Detection Using a Fully Convolutional Network with Compact Polarimetric SAR Images." *Remote Sensing,* vol. 11, no. 18, pp. 2171, 2019.

[4] S. Liu, T. Liu, L. Gao, H. Li, Q. Hu, J. Zhao, and C. Wang. "Convolutional Neural Network and Guided Filtering for SAR Image Denoising." *Remote Sensing* vol. 11, no. 6, 2019.

[5] Liu, X., C. He, Q. Zhang, and M. liao. "Statistical Convolutional Neural Network for Land-Cover Classification From SAR Images." *IEEE Geoscience And Remote Sensing Letters* pp. 1548-1552, 2020.

[6] J. Bromley, J.W. Bentz, L. Bottou, I. Guyon, Y. Lecun, C. Moore, E. Säckinger And R. Shah. "Signature Verification Using a Siamese Time Delay Neural Network". *International Journal Of Pattern Recognition And Artificial* Intelligence, 1993, 7(4): 669-688.

[7] S. Bell. and K. Bala. "Learning visual similarity for product design with convolutional neural networks." *ACM Transactions On Graphics* vol 34, no. 4, pp 1–10, 2015.

[8] D. Robertis, E. M. "Spemann's organizer and self-regulation in amphibian embryos." *Nature Reviews Molecular Cell Biology* vol. 7, no. 4, pp. 296-302, 2006.

[9] A. Gordo. J. Almazan, J. Revaud, and D. Larlus. "End-to-End Learning of Deep Visual Representations for Image Retrieval." *International Journal Of Computer Vision*, vol. 124, no. 2, pp. 237-254, 2017.

[10] S. Ji. S. Wei, and M. Lu. "Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set." *IEEE Transactions On Geoscience And Remote Sensing,* vol. 57, pp. 574-586, 2019.

[11] M. Ma, J. Chen, W. Liu , W. Yang. Ship classification and detection based on CNN using GF-3 SAR images. *Remote Sensing*, vol. 10, pp. 2043, 2018.