Combining Deep Convolutional Neural Network and SVM to SAR Image Target Recognition

Fei Gao
School of Electronic and
Information Engineering
Beihang University
Beijing, China
Email: feigao2000@163.com

Teng Huang
School of Electronic and
Information Engineering
Beihang University
Beijing, China
Dept. of Computer and
electronic engineering
Wuzhou University
Wuzhou, China
Email: huangteng1220@buaa.edu.cn

Jun Wang
School of Electronic and
Information Engineering
Beihang University
Beijing, China
Email: wangj203@buaa.edu.cn

Jinping Sun
School of Electronic and
Information Engineering
Beihang University
Beijing, China
Email: sunjinping@buaa.edu.cn

Erfu Yang
Space Mechatronic Systems
Technology Laboratory,
Department of Design,
Manufacture and Engineering
Management
University of Strathclyde
Glasgow G1 1XJ, UK
Email: erfu.yang@strath.ac.uk

Amir Hussain
Cognitive Signal-Image and Control
Processing Research Laboratory,
School of Natural Sciences
University of Stirling
Stirling FK9 4LA, UK
E-mail: hussain.doctor@gmail.com

Abstract—**Aiming at solving the problem in target recognition of synthetic aperture radar (SAR) images, this paper proposes a novel method based on the combination of Deep Convolutional Neural Network (DCNN) and Support Vector Machine (SVM). The class separation information, which explicitly facilitates intra-class compactness and inter-class separability in the process of learning features, is added to cross-entropy cost function as a regularization term to enhance the feature extraction ability of DCNN. Then the improved DCNN is applied to learn the features of SAR image. Finally, SVM is utilized to map the features into output labels. Experiments are performed on SAR image data in Moving and Stationary Target Acquisition and Recognition (MSTAR) database. The experiment results prove the effectiveness of our method, achieving an average accuracy of 99.15% on ten types of targets.**

Keywords—Synthetic Aperture Radar (SAR); Automatic Target Recognition (ATR); Deep Convolutional Neural Network (DCNN); Support Vector Machine (SVM); Class Separation Information

I. INTRODUCTION

With an increasing amount of data acquired by the SAR imaging system, how to automatically and quickly realize the target recognition of SAR image has become one of the research hotspots.

At present, common methods of Automatic Target Recognition technology of SAR images (SAR-ATR) include template matching [1], Support Vector Machine (SVM) [2], Linear interpolation [3], Principal Component Analysis [4], etc. These methods have been successful in some way, but they rely heavily on experience of experts [5]. Therefore, these methods have a certain blindness and unpredictability when applied to SAR images.

Recently, Deep Learning has been an increasingly hot topic in pattern recognition, and related theories and models have emerged, such as the Deep Restricted Boltzmann [6], Stacked Autoencoders [7] and Deep Convolutional Neural Network (DCNN) [8]. Among them, DCNN has made breakthroughs endlessly in both theory and practice of image target recognition. It can not only automatically extract target features without too much experience of experts, but also deal with the two-dimensional image data directly. These characteristics of DCNN provide a new idea to solve the problems of the automatic feature extraction of SAR targets.

Nowadays, the DCNN-based SAR-ATR system is being gradually proposed. To auto-learn features of SAR images rapidly, general methods start with the framework of AlexNet model. The promotion of training efficiency and recognition accuracy are achieved by optimizing a certain module. For instance, Chen et al. initializes the hyper-parameters of DCNN with an unsupervised sparse autoencoder machine rather than the Backpropagation algorithm used in AlexNet model [9]. The unsupervised sparse autoencoder machine possesses the auto-learning capacity, which accelerates the feature learning of DCNN [10]. The experiment of this method on MSTAR database obtains a target recognition accuracy of 90.1% for 3 classes and 84.7% for 10 classes. Similarly, Li et al. uses the autoencoder machine to initialize the DCNN [11]. The difference is that fully connected layers are used as SNN to work as a final classifier in [11], which greatly reduces training

time of DCNN on the premise of ensuring accuracy. Although the accuracy of two methods is not very high, there is less dependence on experience of experts in the process of feature learning. Some scholars apply DCNN to automatic feature learning as well, whereas the final output layer becomes a SVM classifier rather than a fully connected layer compared with the AlexNet model, constituting DCNN+SVM model [12]. The model makes full use of the DCNN's superiorities in auto-learning of various features and the advantages of the strong generalization ability of SVM [13], avoiding weak representation ability of SVM on high-dimensional samples and poor stability of DCNN. With this framework, the classification accuracy of the method reaches 98.6% [12]. Because of these advantages of DCNN+SVM, such methods have been rapidly developed. Based on the work in [12], Wangner et al. introduce Morphological Component Analysis to preprocess SAR image [14]. It rejects some abnormal testing samples so that the accuracy reaches 99% and the recall comes up to 97.3%.

In summary, the SAR ATR system based on DCNN has achieved varying degrees of success, but it is still in its infancy. The research is generally conducted from the perspectives of the exostructure architecture of DCNN or the Data Augmentation, yet little focus on the optimization of internal functions in DCNN. The CNN+SVM model proposed in [12], for instance, uses the quadratic function as error cost function. Although the classification accuracy is satisfactory, it would be time-consuming if the neurons make an obvious mistake during the training of DCNN. Compared with error cost function, the Cross-Entropy is applied as the quadratic function in [9], [11], [14], while there is a lack of in-depth optimization. To improve the classification ability of CNN, we attempt to add class separability information to cross-entropy cost function as a regularization term in DCNN+SVM model.

The remainder of this paper is organized as follows: Section II describes our DCNN+SVM in detail. Experimental results on the MSTAR database are presented in Section III. Section IV concludes our work.

II. IMPROVEMENT OF DCNN

A. The Introduction of Class Separability Information

To enhance the class separability of the features extracted by DCNN model, the class separability information is added to cross-entropy cost function as a regularization term to train the DCNN. The class separability information consists of intra-class compactness and inter-class separability, which are expressed as E_1 and E_2 respectively:

$$E_1 = \frac{1}{2}\|y_c^n - M_c\|_2^2 \qquad (1)$$

$$E_2 = \frac{1}{2}\sum_{c'}\|M_c - M_{c'}\|_2^2 \qquad (2)$$

where y_c^n denotes the actual output value of the nth training sample which belongs to the cth class. M_c and $M_{c'}$ are the output average values of the training samples in c^{th} and $(c')^{th}$ classes. E_1 and E_2 respectively denote the intra-class distance and the inter-class distance of the output features. Shortening the intra-class distance and increasing the inter-class distance in every iteration are necessary to enhance the separability of

the output features. However, the class separability information can't be introduced to the cost function directly, so it must be converted to a regularization term. Thus, the modified cost function is as follows:

$$E = E_0 + \alpha E_1 - \beta E_2 \qquad (3)$$

where α and β are both weight parameters.

The purpose of modifying the cost function is to adjust convolution kernel and bias matrix to make the network develop in favor of classification, where the error vector is essential.

The error vector of E_1 in the output layer L is:

$$\begin{aligned}\delta_1^L &= \frac{\partial E_1}{\partial z^L} \\ &= \frac{\partial}{2\partial z^L}\|y_c^n - M_c\|_2^2 \\ &= (1 - \frac{1}{n^c})\sigma'(z^L) \circ (y_c^n - M_c) \qquad (4)\end{aligned}$$

where n^c is the number of samples that belong to the class c, σ is nonlinear activation function, z^L is weighted input of L layer, and the symbolic $\circ$ is Hadamard product (or Schur product) which denotes the elementwise product of the two vectors.

The error vector of E_2 in the output layer L is:

$$\begin{aligned}\delta_2^L &= \frac{\partial E_2}{\partial z^L} \\ &= \frac{\partial}{2\partial z^L}\|M_c - M_{c'}\|_2^2 \\ &= \frac{1}{n^c}\sigma'(z^L) \circ \sum_{c'}(M_c - M_{c'}) \qquad (5)\end{aligned}$$

According to (3), (4), and (5), the new error vector of E in the output layer L is:

$$\begin{aligned}\delta' &= \delta^L + \alpha\delta_1^L - \beta\delta_2^L \\ &= \sigma'(z^L) \circ (y_c^n - t^n) + \alpha\left(1 - \frac{1}{n^c}\right)\sigma'(z^L) \circ (y_c^n - M_c) \\ &\quad -\beta\frac{1}{n_c}\sigma'(z^L) \circ \sum_{c'}(M_c - M_{c'}) \qquad (6)\end{aligned}$$

where t^n is the targeted value corresponding to the nth neuron of output layer L.

After getting the error vector in the output layer L, we can calculate the error vector in each layer iteratively by using the Backpropagation algorithm. And the updated parameters of convolution kernel and bias matrix for each layer can be derived by using Gradient Descent algorithm afterwards.

B. Modified Model

A novel DCNN+SVM model of SAR image target recognition is proposed based on above analysis. Firstly, DCNN is trained combining with the Softmax classifier, where the Cross-Entropy added with the class separability information is used as the cost function. When the training of DCNN is complete, Softmax is removed and the top features of DCNN are utilized to train SVM. Finally, the proposed framework DCNN+SVM is constructed, where DCNN is used

to extract sample features and SVM is used as the classifier, as is shown in Fig. 1.
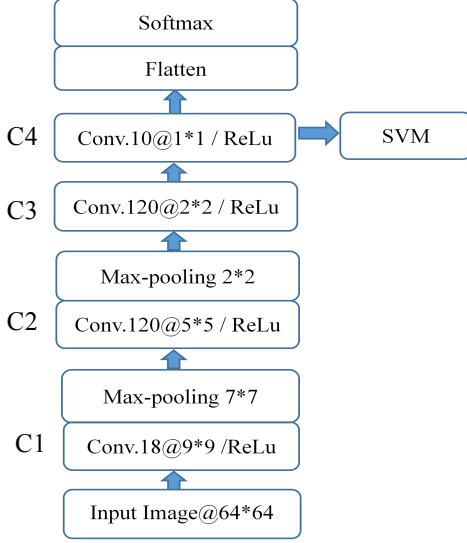

Fig. 1. The proposed DCNN+SVM model

III. EXPERIMENT

A. Experiment Data

To verify the validity of the proposed method, this paper uses the data from the MSTAR database, co-funded by National Defense Research Planning Bureau (DARPA) and the U.S. Air Force Research Laboratory (AFRL).

The MSTAR database includes different types of targets and other related targets under various conditions. This paper applies ten types of targets: 2S1, ZSU234, BMP2, BRDM2, BTR60, BTR70, D7, ZIL131, T62, and T72. As for their high-level classifications, 2S1 and ZSU234 are artilleries; BMP2, BRDM2, BTR60, BTR70, D7, and ZIL131 are assigned to trucks; T62 and T72 belong to tanks. Their SAR images and corresponding optical images are shown in Fig. 2.
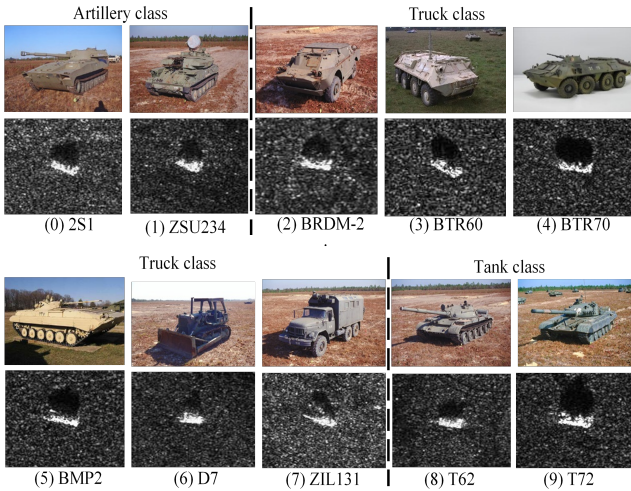

Fig. 2. SAR images and corresponding optical images of ten types of targets in the MSTAR database

The experiment is performed on the above ten types of targets, where the targets under 17^0 depression are taken as the training samples while the ones under 15^0 depression are used as testing samples. The detailed information of type, size, and amount is shown in Table 1. To reduce the input dimension, 64 * 64 pixels in the central part of the patch is intercepted as the input sample under the premise that the complete target still locates at the central position. Then, by shifting and rotating which belong to Data Augmentation technique, each type of training samples is extended to 3000. The Correct Class Probability (p_{cc}) is applied as the evaluation index of our experiment.

TABLE 1   THE TRAINING AND TESTING SET OF OUR EXPERIMENT

Type	Tops	Training Set		Testing Set	
		Depression	No.	Depression	No.
2S1	artillery	17°	3000	15°	274
ZSU234		17°	3000	15°	274
BRDM2	truck	17°	3000	15°	274
BTR60		17°	3000	15°	195
BMP2		17°	3000	15°	195
BTR70		17°	3000	15°	196
D7		17°	3000	15°	274
ZIL131		17°	3000	15°	274
T62	tank	17°	3000	15°	273
T72		17°	3000	15°	274
		sum : 30000		sum : 2503	

B. Our Model Initializes

The parameters of our model are set as shown in Fig. 1. This model includes the input layer, the output layer, 4 convolutional layers and 2 max-pooling layers. Generally, a convolutional layer and a pooling layer are together regarded as a convolutional layer, so there are 6 layers in total. The first convolutional layer C1 (other layers follow the labeling rule) contains 18 feature maps, and the size of convolution kernel and max-pooling are set to 9 * 9 and 7 * 7 respectively. After processing by convolution and max-pooling, the 8 * 8 feature map is outputted. For C2, the size of its convolution kernel is 5 * 5 and the number of feature maps is 120, and a 2 * 2 feature map is outputted after processing by the 2 * 2 max-pooling layer. Therefore, the size of the final feature map is 1 * 1. Besides, ReLu is used as the activation function, and the initial bias and the learning rate are set to 0 and 1 respectively.

C. Experiment Result

The experimental results of the proposed method are shown in Table 2. Table 3 gives the results of contrast method, which applies the standardized quadratic cost function in the training process combining with Softmax and uses Softmax as a final classifier [9]. The average accuracy of our method is 99.15%, 14.45% higher than that of DCNN+Softamx. As for inter-class error, we find no any errors between high-level classes from Table 2. The intra-class confusion in Table 2 is overall satisfactory, with no confusions in tanks and artilleries, only one in trucks. The confusions of BTR70 and BTR60 are relatively more, which may be because BTR70 is an upgraded version of BTR60. For DCNN+Softmax model, however, there are serious confusions in both inter and intra classifications. It proves that adding the class separability information to Cross Entropy error cost function is conducive for distinguishing the high-level classes.

TABLE 2   THE EXPERIMENTAL RESULT OF OUR METHOD

	Artillery		Truck						Tank		
Type	2S1	ZSU234	BRDM2	BTR60	BTR70	BMP2	D7	ZIL131	T62	T72	Pcc(%)
2S1	274	0	0	0	0	0	0	0	0	0	100
ZSU234	0	274	0	0	0	0	0	0	0	0	100
BRDM2	0	0	269	0	0	5	0	0	0	0	98.18
BTR60	0	0	5	188	2	0	0	0	0	0	96.42
BTR70	0	0	1	1	193	1	0	0	0	0	98.47
BMP2	0	0	0	1	2	192	0	0	0	0	98.46
D7	0	0	0	0	0	0	274	0	0	0	100
ZIL131	0	0	0	0	0	0	0	274	0	0	100
T62	0	0	0	0	0	0	0	0	273	0	100
T72	0	0	0	0	0	0	0	0	0	274	100
Total											99.15

TABLE 3   THE EXPERIMENTAL RESULTS WITH THE METHOD IN LITERATURE [9]

	Artillery		Truck						Tank		
Type	2S1	ZSU234	BRDM2	BTR60	BTR70	BMP2	D7	ZIL131	T62	T72	Pcc(%)
2S1	190	1	9	5	5	14	0	21	7	22	69.3
ZSU234	1	249	1	3	0	1	4	6	2	7	90.8
BRDM2	3	0	220	6	18	9	0	15	1	2	80.2
BTR60	4	0	11	168	4	0	4	1	1	2	86.1
BTR70	4	0	4	3	181	3	0	1	0	0	92.3
BMP2	4	4	9	2	9	157	0	6	0	4	80.5
D7	0	7	0	0	0	0	252	5	8	2	91.9
ZIL131	12	0	6	5	7	5	1	226	3	9	82.4
T62	7	2	1	5	0	2	4	7	242	3	88.6
T72	8	1	3	1	1	3	0	9	2	168	85.7
Total											84.7

IV. CONCLUSION

This paper presents a novel method of SAR image target recognition based on CNN. The CNN is improved by adding the class separability information into Cross Entropy cost function and applying Support Vector Machine (SVM) instead of Softmax classifier. The experimental results show that the proposed method achieves a recognition accuracy of 99.15% for ten types of targets with extended training data. Therefore, it proves that our method is effective.

REFERENCES

[1] G. J. Owirka, S. M. Verbout, and L. M. Novak, "Template-based SAR ATR performance using different image enhancement techniques," Algorithms for Synthetic Aperture Radar Imagery Vi, vol. 3721, pp. 302-319, 1999.

[2] Q. Zhao, and J. C. Principe, "Support vector machines for SAR automatic target recognition," Ieee Transactions on Aerospace and Electronic Systems, vol. 37, no. 2, pp. 643-654, Apr, 2001.

[3] J. Ren, J. Jiang, and T. Vlachos, "High-accuracy sub-pixel motion estimation from noisy images in Fourier domain," IEEE Trans Image Process, vol. 19, no. 5, pp. 1379-84, May, 2010.

[4] J. Zabalza, J. C. Ren, M. Q. Yang, Y. Zhang, J. Wang, S. Marshall, and J. W. Han, "Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," Isprs Journal of Photogrammetry and Remote Sensing, vol. 93, pp. 112-122, Jul, 2014.

[5] H. Junwei, Z. Dingwen, H. Xintao, G. Lei, R. Jinchang, and W. Feng, "Background Prior-Based Salient Object Detection via Deep Reconstruction Residual," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 8, pp. 1309-1321, 2015.

[6] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 6, pp. 3325-3337, 2015.

[7] J. Zabalza, J. C. Ren, J. B. Zheng, H. M. Zhao, C. M. Qing, Z. J. Yang, P. J. Du, and S. Marshall, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," Neurocomputing, vol. 185, pp. 1-10, Apr 12, 2016.

[8] M. J. Sun, D. Zhang, J. C. Ren, Z. Wang, and J. S. Jin, "Brushstroke Based Sparse Hybrid Convolutional Neural Networks for Author Classification of Chinese Ink-Wash Paintings," 2015 Ieee International Conference on Image Processing (Icip), pp. 626-630, 2015.

[9] H. W. Sizhe Chen, "SAR target recognition based on deep learning," Data Science and Advanced Analytics (DSAA), 2014 International Conference on, 2014.

[10] J. Fieberg, M. Alexander, S. Tse, and K. St Clair, "Abundance estimation with sightability data: a Bayesian data augmentation approach," Methods in Ecology and Evolution, vol. 4, no. 9, Sep, 2013.

[11] X. Li, C. S. Li, C. P. B. Wang, Z. R. Men, and H. P. Xu, "SAR ATR Based on Dividing CNN into CAE and SNN," 2015 Ieee 5th Asia-Pacific Conference on Synthetic Aperture Radar (Apsar), pp. 676-679, 2015.

[12] S. Wagner, "Combination of convolutional feature extraction and support vector machines for radar ATR," International Conference on Information Fusion, 2014:1-6.

[13] J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel Two-Dimensional Singular Spectrum Analysis for Effective Feature Extraction and Data Classification in Hyperspectral Imaging," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 8, pp. 4418-4433, 2015.

[14] S. Wagner, "Morphological Component Analysis in SAR images to improve the generalization of ATR systems," in Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), 2015 3rd International Workshop on, 2015.