Combination of Multi-Scale Convolutional Networks and SVM for SAR ATR

Xiao Xia[1] ,Yunneng Yuan[1]
1. School of Electronic Engineering, Beihang University
Beijing, China
judexia@buaa.edu.cn, yuan203@buaa.edu.cn

***Abstract*—We apply a novel method to the challenge of synthetic aperture radar (SAR) automatic target recognition (ATR) by combining multi-scale Convolutional Networks (ConvNets) and Support Vector Machine (SVM). The multi-scale architecture permits the classifier to receive more features from different levels of ConvNets, and we can train the model more thoroughly. The application of SVM excels the original Softmax classifier on the nonlinear classification tasks. We conduct our experiments on the Moving and Stationary Target Acquisition and Recognition (MSTAR) database. Average classification accuracy in our experiments can achieve 99.42% on ten-class targets.**

Keywords—synthetic aperture radar; automatic target recognition; convolutional neural network; multi-scale features; support vector machine

I. INTRODUCTION

SAR is one of the most essential earth observation methods at present days. It can provide high-resolution images in any weather, no matter day or night. Due to speckle noises and the special imaging technique of SAR, to recognize a small target in enormous SAR images could be time consuming and usually inoperable. Hence, efficient SAR ATR methods are required desperately.

With the rapid development of ConvNets, we have witnessed plenty of excellent results in different applications such as face recognition [1], handwritten digits [2], traffic sign [3], especially after the critical break through, carried out by Alex Krizhevsky et al. with AlexNet in the ImageNet LSVRC-2010 contest. Abundant researches in SAR ATR using ConvNets have been proposed recent years. For example, Wilmanski et al. analyzed different training approaches (SGD with momentum and weight decay, AdaGrad and AdaDelta) for ConvNets applied to SAR ATR problem [4]. Chen et al. designed a new all-convolutional network. Instead of using the traditional fully connected layer, they replaced it with a convolutional one, which could reduce quantities of trainable parameters, preventing the model from awfully overfitting [5]. Wagner et al. introduced the combination of ConvNets and SVM, taking advantage of the powerful feature extraction ability of ConvNets as well as the excellent generalization ability of SVM [6].

Even though SAR ATR methods based on ConvNets have obtained so many academy's successes, some of which are even state-of-art results, few researches have sought for the refinement of internal structures of ConvNets. In this paper, we extract the multi-scale features from different levels of the network to improve the model's performance.

The structure of the paper is organized as follows. Section 2 describes the architecture improvement of our model. Section 3 presents the experiments on the MSTAR database. Finally, Section 4 is the conclusion of our work.

II. ARCHITECTURE

The differences between the model proposed and other common ConvNets model is the use of multi-scale features and SVM classifier.

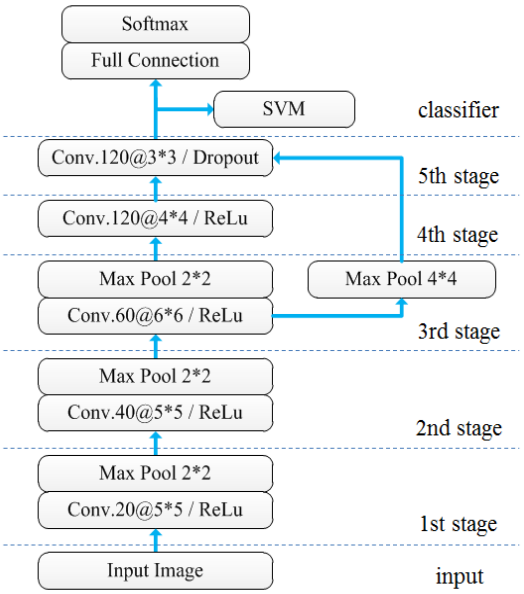

Fig. 1. Multi-scale ConvNets with SVM. The convolutional layers are presented as "Conv. (number of convolutional kernels) @ (kernel size)."

A. *Multi-Scale Features*

While standard ConvNets are designed in strict feed-forward layered structures, which means the output of the former layer could only be fed to the latter one, multi-scale ConvNets combine features from multiple stages in the network by skipping a few middle layers, connecting the lower-level layer to the higher-level layer directly [3]. In our practice, as is shown in Fig. 1, we branched the output of convolutional layer in the 3rd stage by using different max pooling size. The output of the max-pooling layer with 2×2

pooling size is fed to the 4th convolutional layer, whose output is connected to the 5th convolutional layer. In addition to the feature maps obtained by the 4th convolutional layer, we directly connect to the 5th convolutional layer the feature maps yielded by the other max pooling layer in the 3rd stage, with a 4×4 pooling size. Therefore, instead of learning solely the global and invariant features in the 4th stage, the final convolutional layer could receive the features from the former stage, which extracts more local details.

Besides, while training the network with back-propagation algorithm, the gradient would decrease dramatically along with the increase of propagation depth, resulting in a terribly slow weight-updating rate among neurons in the lower-level layers. However, the layer-bypassing connection between the 3rd stage and the 5th stage in our method could reduce the potential gradient loss in the 4th stage to some extent, so we can train the model more thoroughly.

B. Advanced Classifier

Typically, Softmax regression model is applied as the final classifier to the ConvNets. Yet occasionally the full connection strategy in Softmax regression model cannot deal with nonlinear classification tasks perfectly.

In our method, we train the ConvNet with Softmax classifier at first. When the training has been completed, the 120 dimensional features yielded in the 5*th* stage would be utilized for the training of SVM, as is shown in Fig. 1. SVM was first proposed by Vapnik et al. to solve the nonlinear classification tasks by mapping the original lower-dimensional space into a much higher-dimensional space where the features become linearly separable, which could excel Softmax classifier facing nonlinear features extracted by ConvNets.

III. EXPERIMENTS

A. Data Preparation

The SAR image data that used in our experiments were collected by the Sandia National Laboratory SAR sensor platform. The collection was jointly sponsored by National Defense Research Planning Bureau and the Air Force Research Laboratory as part of the MSTAR program [7], using an X-band SAR sensor, in a 1-ft resolution with spotlight mode.

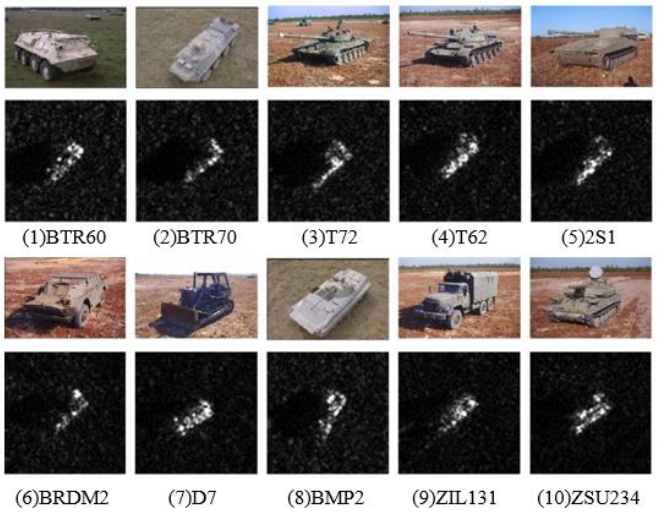

Fig. 2. Optical images and corresponding SAR images of ten categories of targets in MSTAR data sets

The publicly released MSTAR database contains ten different categories of ground targets: BMP2, BRDM2, BTR60 and BTR70 are armored personnel carriers; T62 and T72 are tanks; ZIL131 is truck; D7 is bulldozer; 2S1 is rocket launcher while ZSU234 is air defense unit. The optical images and corresponding SAR images of the ten categories of targets are shown in Fig. 2.

In our experiments, we take the targets images under $17°$ depression angle as training sets while the targets images under $15°$ depression angle as testing sets. Numbers of images in the training and testing sets are shown in TABLE I. .

TABLE I. NUMBER OF IMAGES IN THE TRAINING AND TESTING SETS

Class	Train		Test	
	Depression	No.	Depression	No.
BMP2	17°	233	15°	195
BTR70	17°	233	15°	196
T72	17°	232	15°	196
BTR60	17°	256	15°	195
2S1	17°	299	15°	274
BRDM2	17°	298	15°	274
D7	17°	299	15°	274
T62	17°	299	15°	273
ZIL131	17°	299	15°	274
ZSU234	17°	299	15°	274

Before fed to the ConvNet, the data sets are augmented by image translations and rotations. We employed this by conducting a small random angle rotation to the original images and then randomly extracting patches from the rotated images. Depending on this approach, we expand our training sets to 3000 images per category while the input dimensions have been reduced in the mean time, both of which could help to reduce overfitting.

B. Configuration Specifics

As is demonstrated in Fig. 1, this multi-scale ConvNet model contains an input layer, an out layer, five convolutional layers and four max-pooling layers. In addition, we choose ReLU as activation function for every convolutional layer, and we set the convolution stride to one pixel.

Filtered by 20 convolution filters of size 5×5 in the 1st convolution layer, the 80×80 input image would yield 20 feature maps of size 76×76, whose size, after max pooling, becomes 38×38. The 2nd stage contains 40 convolution filters of size 5×5, and yielding 17×17 feature maps after the max-pooling layer. In the 3rd stage, after 60 feature maps of size

TABLE II. THE CONFUSION MATRIX OF OUR METHOD

Class	BMP2	BTR70	T72	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU234	P_{cc}(%)
BMP2	195	0	0	0	0	0	0	0	0	0	100
BTR70	0	195	0	1	0	0	0	0	0	0	99.49
T72	0	0	196	0	0	0	0	0	0	0	100
BTR60	0	0	1	189	0	2	0	0	2	1	96.92
2S1	0	0	0	0	272	1	0	1	0	0	99.27
BRDM2	1	0	0	0	0	272	0	0	1	0	99.27
D7	0	0	0	0	0	0	273	0	1	0	99.64
T62	0	0	1	0	0	0	0	272	0	0	99.63
ZIL131	0	0	0	0	0	0	0	0	274	0	100
ZSU234	0	0	0	0	0	0	1	0	0	273	99.64
Total											99.42

12×12 are outputted, we apply two different max-pooling operations with different size (2×2 and 4×4) to these feature maps. The max-pooling layer of size 4×4 yields 60 feature maps with the size of 3×3, while the outputs of the other max-pooling layer are passed to the 4th convolutional layer, the outputs of whom are 120 feature maps with the same size of 3×3. Then both the 60 feature maps and the 120 feature maps are concatenated and fed into the 5th convolutional layer, who has 120 convolutional filters of size 3×3, yielding 120 feature maps of size 1×1.

We initialize the network with zero-mean Gaussian distributions with a standard deviation of $\sqrt{2/n}$, where n denotes the number of inputs of each unit [8]. Then the network is trained by mini-batch stochastic gradient descent with a batch size of three images, along with a momentum parameter of 0.9 and a weight decay parameter of 0.0005. Gradients are calculated by back-propagation. Learning rate is set to 0.001 initially and is multiplied by a factor of 0.5 after every 20 epochs. The parameters above are obtained through cross validation.

We choose radial basis function as kernel function for SVM, and seek out the parameters *C* and *gamma* by grid search and 3-fold cross validation.

C. Experimental Results

The confusion matrix of our experiment with the proposed multi-scale ConvNet with SVM is shown in TABLE II. . Each row in the table denotes the actual target class while each column in the table denotes the predicted class by the network. The average ten-class target recognition accuracy with our method is 99.42%.

To examine the improvements that multi-scale features bring to the classification tasks, we train a regular ConvNet whose structure is the same as our proposed multi-scale ConvNet except for we remove the layer-bypassing connection between the 3rd stage and the 5th stage. In this experiment, we

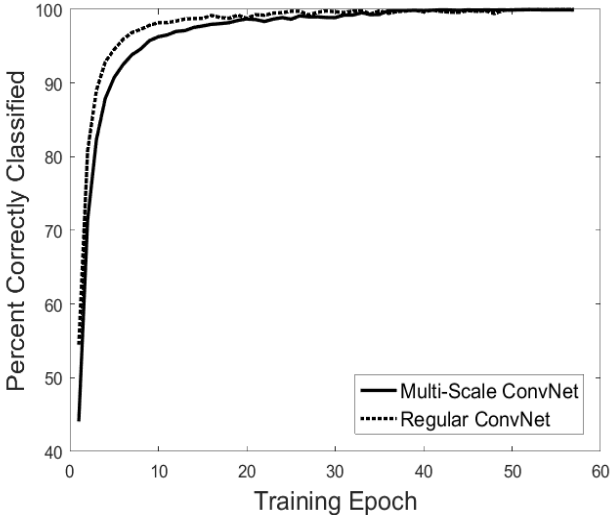

Fig. 3. The training classification accuracy versus training epoch

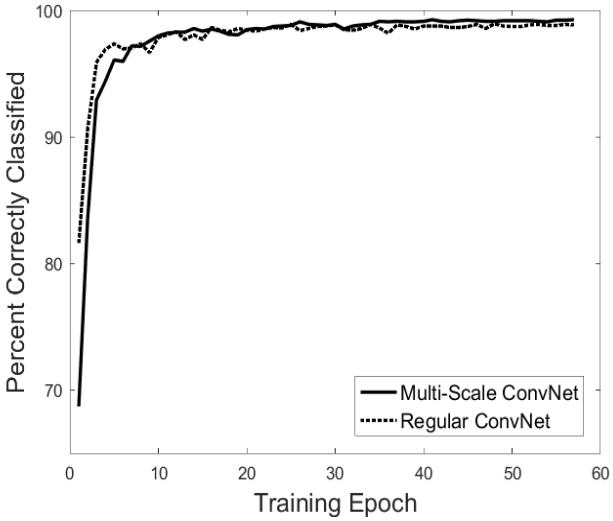

Fig. 4. The testing classification accuracy versus training epoch

use Softmax classifier instead of SVM classifier for convenient conducting. We compare both the training classification accuracy and the testing classification accuracy versus training epoch of these two ConvNets, as is shown in Fig. 3 and Fig. 4.

Due to extra parameters the layer-bypassing connection brings to the ConvNet, we notice that the training speed of multi-scale ConvNet is slower than the regular ConvNet. Moreover, extra parameters usually bring to ConvNets overfitting, which means the classification performance on the testing sets would get worse. According to Fig. 4, however, the performance of multi-scale ConvNet on the testing sets excels the regular ConvNet, reaching an average classification accuracy of 99.26% , while the average classification accuracy of regular ConvNet could only reach 99.09% . Therefore, we can clearly conclude the multi-scale features could improve the performance of ConvNets, owning to the fact that they can feed the classifier with more features and can make the training procedure more thoroughly.

TABLE III. RESULTS COMPARED WITH OTHER METHODS

Method	Accuracy (%)
SVM [9]	90
AdaBoost [9]	92
A-ConvNet [5]	99.13
ConvNet with SVM [6]	98.16
our method	99.42

The experimental results of our method are also superior to the other methods listed in TABLE III. .

IV. CONCLUSIONS

In this paper, we propose a novel method for SAR ATR problem. In order to reduce the potential gradient loss, utilize features from different levels and better solve the nonlinear classification tasks, we introduce multi-scale features to our network and replace the Softmax classifier with SVM. The average ten-class target recognition accuracy on the MSTAR database is 99.42% . The experimental results demonstrate the effectiveness of our method.

REFERENCES

[1] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891-1898.

[2] Y. LeCun, L. Bottou, Y. Bengio, et al, "Gradient-based learning applied to document recognition," in *Proc. IEEE 86(11)*, 1998, pp. 2278-2324.

[3] P. Sermanet, Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," *Neural Networks (IJCNN), The 2011 International Joint Conference on. IEEE*, 2011, pp. 2809-2813.

[4] M. Wilmanski, C. Kreucher, and J. Lauer, "Modern approaches in deep learning for SAR ATR," *SPIE Defense+ Security. International Society for Optics and Photonics*, 2016, pp. 98430N-98430N.

[5] S. Chen, H. Wang, F. Xu, et al, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing 54.8*, 2016, pp. 4806-4817.

[6] S. Wagner, "Combination of convolutional feature extraction and support vector machines for radar ATR," *Information Fusion (FUSION), 2014 17th International Conference on. IEEE*, 2014, pp. 1-6.

[7] E.R. Keydel, S.W. Lee, and J.T. Moore, "MSTAR extended operating conditions: A tutorial," *Aerospace/Defense Sensing and Controls. International Society for Optics and Photonics*, 1996, pp. 228-242.

[8] K.M. He, et al, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE International Conference on Computer Vision*, 2015, pp. 1026-1034.

[9] U. Srinivas, V. Monga, and R.G. Raj, "SAR automatic target recognition using discriminative graphical models," IEEE Transactions on Aerospace and Electronic Systems 50.1, 2014, pp. 591-606.

.