Hindawi
Journal of Sensors
Volume 2019, Article ID 1246548, 9 pages
https://doi.org/10.1155/2019/1246548

Research Article

A Novel Convolutional Neural Network Architecture for SAR Target Recognition

Yinjie Xie,[1] **Wenxin Dai**,[1] **Zhenxin Hu**,[1] **Yijing Liu**,[1] **Chuan Li**,[1] **and Xuemei Pu**[2]

[1]*College of Computer Science, Sichuan University, Chengdu 610065, China*
[2]*College of Chemistry, Sichuan University, Chengdu 610065, China*

Correspondence should be addressed to Chuan Li; lcharles@scu.edu.cn and Xuemei Pu; xmpuscu@scu.edu.cn

Received 10 January 2019; Revised 20 March 2019; Accepted 24 March 2019; Published 5 May 2019

Guest Editor: Hyung-Sup Jung

Among many improved convolutional neural network (CNN) architectures in the optical image classification, only a few were applied in synthetic aperture radar (SAR) automatic target recognition (ATR). One main reason is that direct transfer of these advanced architectures for the optical images to the SAR images easily yields overfitting due to its limited data set and less features relative to the optical images. Thus, based on the characteristics of the SAR image, we proposed a novel deep convolutional neural network architecture named umbrella. Its framework consists of two alternate CNN-layer blocks. One block is a fusion of six 3-layer paths, which is used to extract diverse level features from different convolution layers. The other block is composed of convolution layers and pooling layers are mainly utilized to reduce dimensions and extract hierarchical feature information. The combination of the two blocks could extract rich features from different spatial scale and simultaneously alleviate overfitting. The performance of the umbrella model was validated by the Moving and Stationary Target Acquisition and Recognition (MSTAR) benchmark data set. This architecture could achieve higher than 99% accuracy for the classification of 10-class targets and higher than 96% accuracy for the classification of 8 variants of the T72 tank, even in the case of diverse positions located by targets. The accuracy of our umbrella is superior to the current networks applied in the classification of MSTAR. The result shows that the umbrella architecture possesses a very robust generalization capability and will be potential for SAR-ART.

1. Introduction

Synthetic Aperture Radar (SAR) could provide very high resolution images in all-weather day-and-night conditions [1]. Thus, it has been widely applied in national economy and military fields [2, 3]. Unlike optical image with rich colors, the SAR images are characterized by the strength of the pixel grayscale, in which the regions with high intensity represent targets. The pixel value is mainly derived from two kinds of reflection of the electromagnetic waves. The first is the single reflection from the surface of the target, which depends on the surface roughness, the shape, and the material of the target. The second is secondary reflection of the electromagnetic waves upon the dihedral corner between the target and the ground, which has a great connection with the height of the radar and the shooting angle. Overall, SAR has the characteristics of scattering electromagnetic, high resolution, speckle noise [4], huge size, and single channel. These characteristics make the SAR image data information large and the target electromagnetic scatter complicated, in turn taking a lot of manual works to recognize targets in the massive SAR images. Thus, the SAR Automatic Target Recognition (ATR) is challenging and becomes one of the research hotspots for remote sensing technology. A general architecture of SAR ATR is composed of three parts: detection, discrimination, and classification [5]. The detection is to extract target regions by a constant false alarm rate (CFAR) detector [6]. Then, the discriminator is used to identify these candidate regions located by targets according to the output of the first stage. At the final stage, a classifier is utilized to recognize the category of each target type.

Current mainstream classification methods of SAR-ATR generally include three types: template-based method [7], model-based method [8], and pattern-based method [9]. The

typical template-based SAR ATR system utilizes minimum mean square error (MSE) criteria to identify the target type from a database of stored target reference images or templates [7, 10]. The model-based system analyzes each image in detail and identifies each part of a signature contribution to recognition [8, 11]. Compared with the two methods, the strategy based on the pattern recognition devoted great contribution to the image classification in the past several years. In general, the pattern-based architecture first designs a set of feature extractors to convert the raw image into low-dimensional feature vectors and then the output vectors are categorized by a classifier. Some ART algorithms have been widely applied to the SAR image classification and recognition, for example, artificial neural networks [12], support vector machines [13], and convolutional neural networks [14]. In particular, deep learning based on convolutional neural networks (CNN) has been considered to be one of the most comprehensive methods in the SAR image classification and detection.

However, due to the limited data set of SAR images [15], the SAR ATR task using the convolutional neural network easily causes overfitting. To address this problem, three main strategies were employed. One is to use transfer learning [16], which first pretrain a CNN from a large data set and then fine-tune the network on the specific task for the small SAR data set. The pretraining data set may be selected from large number of labeled optical images. However, due to the difference between the optical and SAR images, it may not perform well for the SAR images. Alternatively, a lot of unlabeled SAR images may be appropriate to take place of the optical images [16]. But, the acquisition of unmarked SAR data set is still difficult and requires a lot of manual works. The third way to overcome the limitation of SAR data set is data augmentation [17]. The SAR image data set used are mostly standard data, in which the target location is usually fixed in the center. As known, in the actual situation, the target location is often random. Therefore, considering translation, speckle noise, and rotation as data augmentation is a good way, which not only could overcome the limitation of the small data set but also simulate the actual situation of the locations of the targets. However, the way is usually not taken into account in most studies. Furthermore, the performance on the results of the works with inclusion of the data augmentation using CNNs was not very high, which should be attributed to that the CNN architecture used in the works should be further improved.

In general, the improvement strategy regarding the CNN architectures involves in the increase of the width and the depth of the network. However, simply increasing the depth size may cause a problem of vanishing/exploding gradients [18, 19]. To solve this problem, residual network (ResNet) [20] is proposed, which is composed of linear superposition of many residual modules. Each module sums the input value to the output value after two-layer convolution, which makes the weight parameter adjustment of the network layer more reasonable with the aid of the theory of identity mapping, and thus could avoid the problem of vanishing/exploding gradients with the increasing depth. Using a deep residual network with more than 100 layers, the error was reduced to be 3.57% for classification task on the ImageNet data set [20]. But, the increase in the depth is not unlimited because too deep network still leads to the problem of overfitting. The other strategy is to increase the width of the CNN architecture so that more features could be utilized. But, simply increasing the width would lead to a large number of parameters and more computational resources, which may also cause the overfitting. In order to address the problem, some techniques like inception [21, 22] and X-ception [23] were introduced to the CNN architecture to optimize its network structure. The inception/X-ception modules do not simply increase the width but divide a number of channels into segments independent. Then the segments with different configurations are the concatenation fusion of the feature extraction from different scales so that enough features could be obtained but computational difficulties could be avoid. Recently, a joint network architecture was proposed through adding an inception module into ResNet (Inception-ResNet) so that it could simultaneously take into account the width and the depth. The experiment on OLSVRC-2012 proved that the Inception-ResNet could significantly accelerate the training of the network and achieve better accuracy than the single inception network [24]. Although these strategies above were demonstrated to improve performance for the optical image classification, unfortunately, they were not applied to the SAR field. Furthermore, the characteristics of the SAR image are different from the optical ones. Thus, it is inappropriate for the methods successfully applied in the optical images to be directly and simply applied to the SAR-ATR field. Further improvement will be needed.

In the work, we proposed a novel CNN architecture suitable for the SAR data through improving the Inception-ResNet architecture from the optical images. In the architecture, we more focused on the extraction of features due to fewer representatives from the SAR image than the optical one. Thus, different from the Inception-ResNet algorithm in the optical classification, we embedded the ReNet module into the Inception one so that the network could extract sufficient features through the inception module and simultaneously utilize the advantage of the ResNet for the network depth. The novel CNN architecture is named umbrella algorithm in the work. It is composed of six separate segments based on inception, each of which has different convolution configurations extracting different levels of features with the aid of the ResNet module. The architecture possesses stronger ability to fuse the feature extracted from different scales than the common inception module. Our experimental results confirm that the umbrella architecture could achieve excellent classification accuracy for 10-class targets of vehicles and eight variants of T72 tanks from the Moving and Stationary Target Acquisition and Recognition (MSTAR) program.

2. Materials and Methods

2.1. Learn about Convolutional Neural Networks (CNNs). Figure 1 shows a typical deep learning network, which generally includes convolutional layers, pooling layers, and fully connected layers. The convolutional layer consists of a number of convolution kernels, which is a two-dimensional

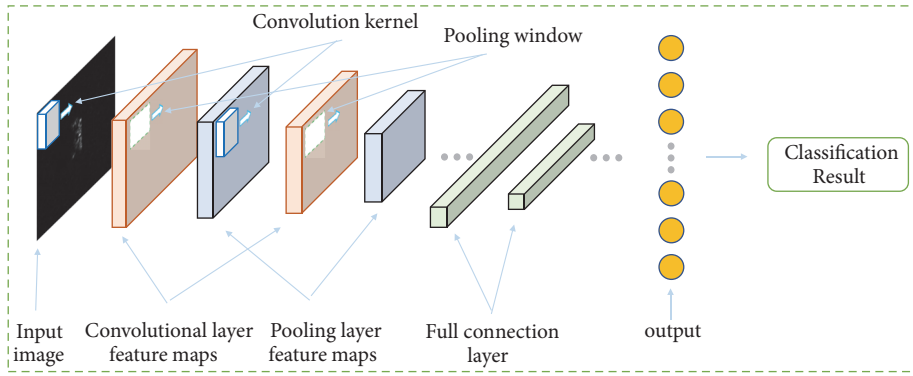

Figure 1: The structure of CNN.

matrix of weights W. The convolution kernel convolves an input image (also a two-dimensional matrix) in the form of a sliding window. Then a matrix called feature map is obtained. Consequently, the convolutional layer implements different channels of feature extraction through multiple convolution kernels. The pooling layer is also called the sampling layer, which could use the sliding window to convolve the input or the feature map so that it could reduce the feature dimension and the amount of calculation. The pooling process is generally implemented using maximum pooling or average pooling, which denotes the maximum value or average value of the selected sliding window. Different from the convolutional layer, the pooled layer does not involve in weights and parameters. In general, each feature map of the input is pooled in the same way and the number of features of the original input remains unchanged. Then, the fully connected layer is used to map the feature representations from the convolutional layer to the sample space in order for classification, which is composed of a group of neurons and connections with weight values. Since the number of the parameters of the fully connected layers is very large, some networks used the convolutional neural networks to take place of the fully connected layers. In order to improve the prediction ability of the CNN model, the following ways were usually utilized. One effective way is to train a large number of data set to learn enough sample feature information so that the test set can be better explained. In addition, a good network architecture model is required, which generally consists of reasonable network layers and network width as well as addition of some other processing ways, for example, adding batch normalization [27] layer, dropout [28] layer, and regularization [29]. Another way is to configure the network's hyperparameters like the number of convolution kernels and the size of the convolution kernel. These parameters are closely associated with the number of the weight W, which directly affects the performance of the model. These parameters are often designed according to some experiences, which also refer to the configurations of some existing models. In a whole, a key is to consider the size of the task when designing a model. This includes the size of the sample set and the size of the sample feature information. When the feature information is relatively rich, the model could achieve a good result by designing a network with sufficient depth.

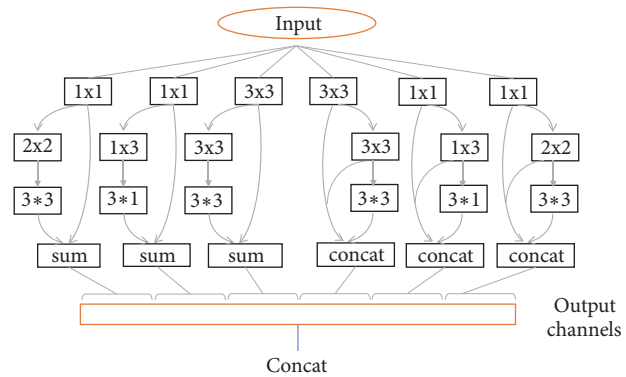

Figure 2: The umbrella module of the basic component of the proposed architecture.

2.2. Module of Umbrella from the Work. Similar to the previous Inception-ResNet network in the optical image, the umbrella module decomposes the input into independent feature mapping channels so that it could learn diverse characteristics of the input space from different levels. However, different from the previous architecture, the ResNet module was embedded into the Inception module in the work, rather than the inception one embedded in the ResNet, considering that the samples and feature information of the SAR image are generally less than those of the optical one. Figure 2 illustrates the structure of one module of umbrella. It contains six different feature extraction paths, which are represented by six brackets with specific roles. Each path contains 3 convolution layers. The six paths were divided into two categories, as shown in Figure 2. The three paths on the left use the residual network [20] to sufficiently extract the input space features, in which the parameters of each layer of the convolution layer can be fully trained so that it could avoid falling into local optimum. The three paths in the right are convolved in terms of the traditional convolution method. But, different from the traditional convolution method, we obtain the feature map of each convolutional layer through concatenating three layers in order to extract more feature information. Finally, the output of the six paths is further concatenated by the feature extracted from different depth.

In order to avoid increasing calculation overhead, the number of convolution kernels of each layer is controlled

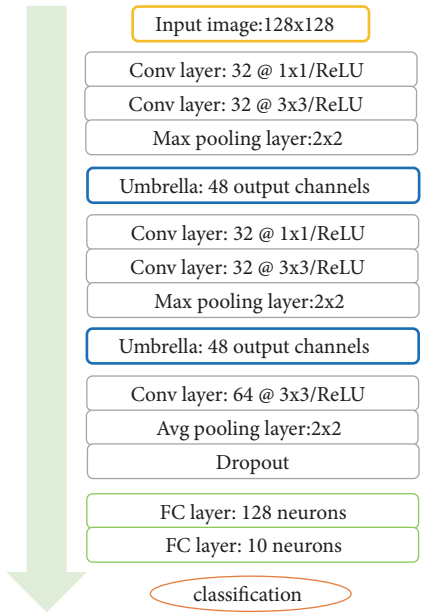

Figure 3: The umbrella architecture of the proposed method.

to be lower than 64, which is enough to learn rich feature information due to the module's independent decoupling of the input feature space. In the first path, we first used 1x1 convolution kernels. It could reduce the parameters and the amount of computation because the convolution kernel requires weight sharing. In addition, the 1x1 convolution kernel could preserves all the information of the input space and increase depth of one layer, which was proved to be useful in deep learning [21, 30]. In order to extracting deeper features, we used the 2x2 and 3x3 convolution kernel in the subsequent step. The second path also first uses the 1x1 convolution kernel. In order to extract different features, we used the 1x3 and 3x1 convolution kernel in the subsequent step, which could reduce parameters with respect to the 2x2 and 3x3 convolutional kernel. The third path uses the conventional convolution kernel with the 3x3 convolution kernel. The right three paths are similar to the left ones. After that, we used the BR layer behind the path to normalize the input values, which enables the activation function to obtain a reasonable feature map from the input space and was already proved to be effective for improving the network [27].

2.3. Methodology. The complete framework of Umbrella model is shown in Figure 3. The network contains a total of 3 convolutional blocks and 2 umbrella blocks, which contain 12 convolutional layers, 2 fully connected layers, and 3 pooling layers. Consequently, the whole architecture of the umbrella model contains a total of 1344 convolution kernels and 131 full connected layer neurons, which lead to a total of 26368 weights. The full connection layer parameters vary according to the input size. For example, if the image is the 128x128 in size, the parameters of the full connection layer are 30080. The convolutional layer is mainly applied to extract the features while the fully connected layer serves as a regression classification. Only one dropout layer is used in the network in order to reduce overfitting.

The network was implemented by Keras [31], which is a high-level neural network API with the backend of Tenser-Flow, CNTK, and Theano. In training, stochastic gradient descent was utilized with a momentum of 0.9, a learning rate of 0.005, and a decay of 0.0004. The experiment was performed on the Linux operating system and an NVIDAI GPU GTX1080Ti.

3. Results and Discussion

The proposed method is verified and discussed by experimental results in this section, which is organized in terms of the four sections. Subsection 3.1 gives a brief introduction of the data set. The main experimental results and discussions are listed in Subsection 3.2, in which our result is also compared with some current state-of-the-art SAR recognition methods. Subsection 3.3 takes into account the case of data augmentation with the aid of the algorithm of noise. In addition, the results from our CNN architecture are also compared with those from some advanced algorithms of the optical image classification in the Subsection 3.4. In Subsection 3.5, our method is further applied to classify the eight variants of T72 tanks, which were not nearly covered by previous SAR-ATR systems due to their high similarities. In the work, the performance evaluation is mainly based on the two criteria. One is the accuracy rate measured in terms of the following equation:

$$A = \frac{N_c}{N_t} \tag{1}$$

It is the ratio of the correct number N_c predicted to the total number N_t in the sample sets. It is one of the most important indexes in the classification and recognition. The other criterion is the loss value, which could measure the loss and error of the true value and the predicted value, as shown by the following[32]:

$$J = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log\left(\hat{y}_n\right) + \left(1 - y_n\right)\log\left(1 - \hat{y}_n\right)\right] \tag{2}$$

where $\hat{y}_n$ denotes the predicted value, y_n denotes the ground truth, and N is the number of samples. Contrary to the accuracy, the lower the loss value, the better the model performance. Therefore, the two parameters are taken together to evaluate the performance of one model.

3.1. SAR Data Set. The experimental data comes from the measured SAR ground static Target announced by the MSTAR (Moving and Stationary Target Acquisition and Recognition) program, supported by the Defense Advanced Research Project Agency (DARPA) and the Air Force Research Laboratory (AFRL) [33]. The sensor that collects the data set is a high-resolution spotlight synthetic aperture radar with a resolution of 0.3m x 0.3m, operating in the x-band and polarization mode of HH polarization. The data set contains 10 types of ground vehicles, such as

Table 1: The number of samples for the 10 class vehicles.

Class	2S1	BMP2	BRDM2	BTR_60	BTR_70	D7	T62	T72	ZIL131	ZSU_234	Total
Train set(17)	200	200	200	200	200	200	200	200	200	200	2000
Test set(15)	274	195	274	196	196	274	273	196	274	274	2425

Table 2: The number of samples for eight T72 variants.

Variants	A04	A05	A07	A10	A32	A62	A63	A64	Total
Train set(17)	299	299	2999	296	298	299	299	299	2388
Test set(15)	274	274	274	271	274	274	274	274	2189

Table 3: Accuracy for umbrella versus state-of-the-art method.

Method	Date	Train	Test	Acc(%)
MSRC [25]	2014	2747(17)	3203(15)	93.66
TSJR [26]	2015	3671(17)	3203(15)	93.41
A-ConvNet [1]	2016	2747(17)	2426(15)	99.13
DCHUN [15]	2017	2000(15)	2462(17)	99.09
CNN-TL [16]	2017	2747(17)	2425(17)	99.09
Umbrella	2017	2000(17)	2425(15)	99.54

The numbers 17 and 15 in parentheses indicate the shooting angle of SAR. The date denotes publication time.

BTR70 (armored transport vehicle), BMP2 (infantry fighting vehicle), T72 (tank), 2S1 (self-propelled howitzer), BRDM2 (armored reconnaissance vehicle), BTR60 (armored transport vehicle), D7 (bulldozer), T62 (tank), ZIL131 (cargo truck), and ZSU234 (self-propelled artillery). In addition, the data set also includes the 8 variants of T72 tanks with different military-configurations, for example, machine guns, fuel tanks, and the antennas. Since the SAR image is very sensitive to the azimuth factor, the images with different orientations were collected, in which the range of orientation is from 0° to 360° at the interval of 1° to 2 °.

In the work, for the 10 types of object targets and the eight variants of T72, we used data with a shooting angle of 17° for the training set and one with 15° for the test set, which have 128x128 resolution. The number of each class under study is listed in the Tables 1 and 2.

Figure 4 representatively shows the optical and the SAR images for the ten vehicles and the eight variants of the T72 tanks. It can be seen that most vehicles have high intensities, except for BMP2, BTR70, and T72. The low intensities of the three vehicles may be attributed to their relatively low heights which lead to a drop in the dihedral corner reflection and special surface materials with respect to the other vehicles which decrease the single reflection. Another possible reason is that the difference in radiation correction between different images makes the overall scattering intensity of these three targets lower than others.

3.2. Experiments on the Standard Data Set of 10 Class Vehicles. As shown in Table 2, the training set contains 2000 samples and the test set includes 2425 samples for the 10 categories of targets. Each category between the training set and the test set is the same for the serial, the configuration, and the version of the target. The difference is only the shooting angle (17° for the train set and 15° for the test set). In addition, each of the images contains only one complete target in its central location. Figure 5 shows the result of the confusion matrix for the test set. The abscissa denotes the predicted label while the ordinate indicates the real label. The digit in the grid of the diagonal position denotes the number of predictions matching the real labels. The other digits in the figure denote the number of targets misclassified.

As shown in Table 3, the accuracy for the classification can achieve 99.54% for the test set, derived from our method. It can be seen from Figure 5 that the maximum number is only three for samples misclassified, in which three BTR_60 vehicles are misidentified as BRDM-2 due to their similarities, as reflected by Figure 4. Even so, the correct recognition rate of this category still reaches 98.90%. Furthermore, BMP2 and BTR70 and T72 with low pixel values also show almost 100% accuracy, further confirming the reliability of our method.

In addition, the performance of our umbrella network is also compared with some recent results from advanced SAR identification methods, including transfer learning based method (CNN-TL-bypass) [16], sparse representation of monogenic signal (MSRC) [25], tritask joint sparse representation (TJSR) [26], A-ConvNet [1], and DCHUN [15]. As evidenced by Table 3, the accuracy rate of our method is improved by 0.41%~0.45% with respect to these methods based on deep learning.

3.3. Experiments on Augmentation Data for the Ten Classes of Vehicles. The result of Section 3.2 only takes into account the situation that all targets of the image set are fixed in the middle position. However, in actual situations, most image acquisitions are not the standard, which do not uniformly constrain all targets to the same position. Thus, it is more practical to consider the case that the positions of the targets on the image should be random. Based on the consideration, we extended the original data set through translation of the

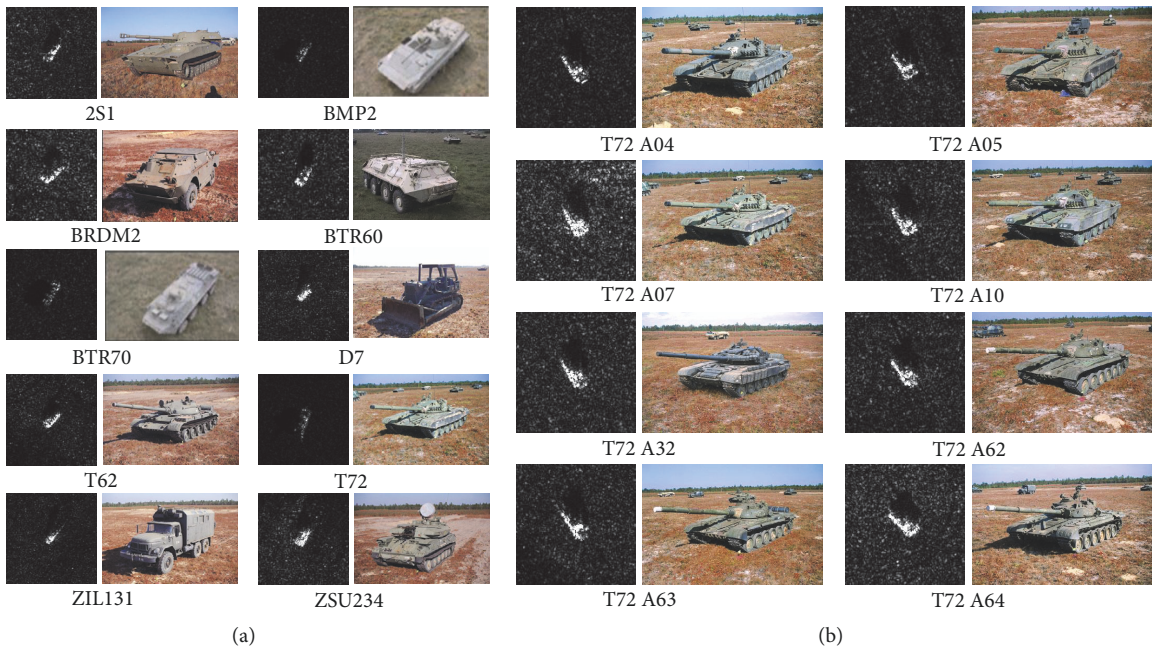

Figure 4: The comparison between optical images and SAR images. (a) Examples on the 10 class vehicle; (b) examples on the 8 variants of T72 tanks.

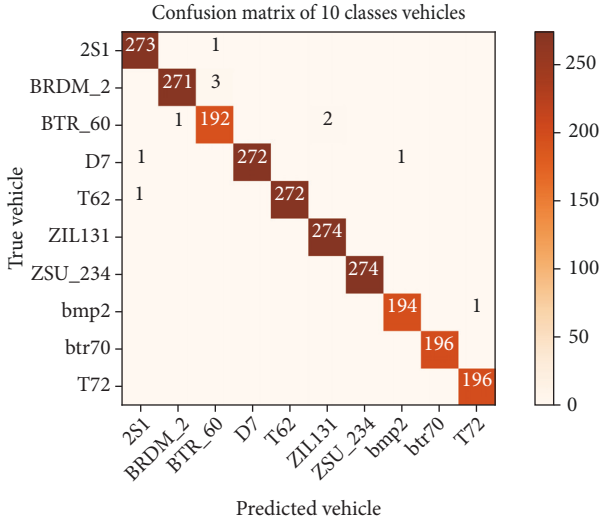

Figure 5: The confusion matrix of 10 classes vehicles. The accuracy of the test set is 99.54%.

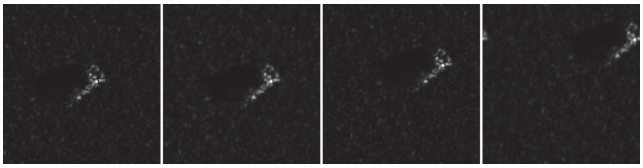

Figure 6: Illustration of random translation. The levels of translation are 10%, 20%, and 30%, respectively.

Table 4: Accuracy across the level of shift.

Shift	10%	20%	30%
Accuracy(%)	31.79	22.39	21.19

image by different levels (10%, 20%, and 30%), as shown in Figure 6. In order to evaluate the necessity of the data extension, we designed and discussed several models. One model is that data of the train set is standard without any translation while ones of the test set were translated by 10%, 20%, and 30%. As shown in Table 4, the accurate rates are very low. The larger the translation-extent, the lower the accurate rate. Not unexpectedly, the features derived from the train set almost focus on the center position located by the target while the targets are deviated from the center position in the test set. Thus, based on the feature from the center position, the model hardly gets accurate identification for the test set.

In other words, the model only considering the center position cannot cope with the actual needs. In order to improve the robustness of the model, we constructed the train set with consideration of the uncertainty of target positions, in which the number of the targets at the original image was extended by 10 times through the random translation manner. The random shift levels involve in 10%, 20%, and 30%. Similarly, the test set was also extent by the same translation. Consequently, the data of the train set and the test set are 10 time larger than the previous ones, as shown in Table 5. It can be seen from Table 5 that the accurate rates are increased from the 31.79% to 99.12%-99.34%.

The result indicates that the generalization ability of the test set is significantly improved when the training set considers the translation. The result verifies the necessity of data extension. In other words, if the test set contains the targets in diverse situations, the train set must learn the information from the samples in the train set. When more

Table 5: The number of samples in the train set and the test one and their accuracies[1].

Data set	Train	Test	Acc(%)
Standard	2000	2425	99.54
Ag of test set	2000	24250	31.79
Ag of train set	36710	2425	99.34
Ag of all targets	20000	24250	99.12

[1]Standard denotes the targets without any translation and Ag denotes the targets in the data set including 10% random shift.

Table 6: The loss and accuracy for umbrella versus other methods.

Method	Loss	Acc(%)
Umbrella	0.027	99.54
Vgg16	0.073	97.23
ResNet50	0.164	95.51
Inception_ResNet	0.047	98.35
Inception_v4	0.109	96.12
Xception[1]	0.096	96.90

[1]When the network was applied to the MSTAR data set, the number of layers of the network was reduced slightly to fit the image size.

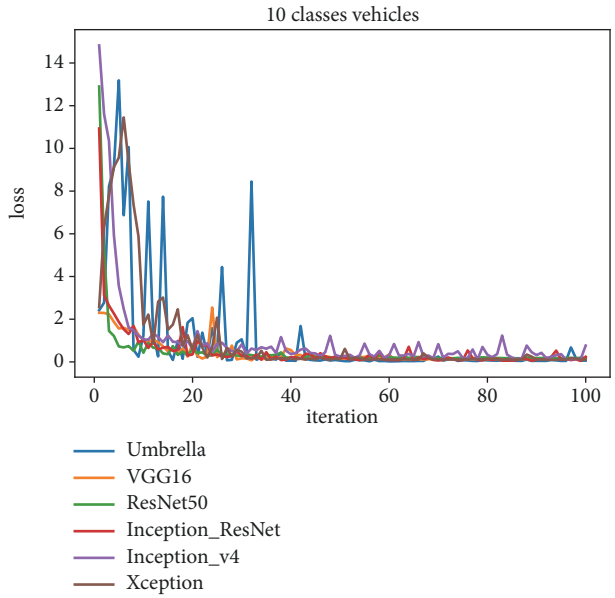

Figure 7: The loss-iteration of different methods in 10 categories classification task.

practical situations are taken into account in the data set, the model is undoubtedly more robust and easier to identify the targets under different scenarios.

3.4. Experiments vs. the State-of-the-Art CNN Architecture. In order to further evaluate the performance of the umbrella method proposed by us, we compared it with some excellent network architectures, based on the same data sets of the ten-class vehicles. These architectures exhibit excellent performance in the optical image but have not been applied to the SAR image. The performance comparison of these methods is shown in Figure 7 and Table 6.

As reflected by Figure 7, the loss value is decreased with increasing iterations and that from our umbrella is lowest. It can be seen that the Umbrella model also presents the best accuracy, compared with the others. Although these deep convolutional neural networks applied in the optical images also took into account increasing the depth and width in order to improve their performances, they do not improve the performance of the recognition for the SAR images with respect to our umbrella method. As mentioned above, the optical images are rich in color information and have distinct target characteristics while the SAR images are displayed by the grayscale values of different intensities with a small target. Therefore, the direct application of these methods with high performance in the optical images cannot achieve the same performance. The result further indicates the necessity of constructing one new architecture appropriate for the SAR image like our umbrella.

3.5. Experiments on the Standard Data Set of Eight Types of T72 Tanks. The similarity of the eight types of T72 tanks is much higher than that of the ten types of vehicles. In order to further assess the performance of the umbrella model, it is further applied to identify the eight types of T72 tanks, which was nearly not taken into account in previous SAR-ART works. As one of few, Dr. Du [34] used a CNN network to classify the eight variants of T72, achieving 94.8% accuracy. Figure 8 shows the classification results of the T72 variants.

It can be seen that the number of targets matching real labels is large and only few samples in every classes were wrongly predicted, not more than 15. The recognition rate of the umbrella model still achieves 96.35%. The result is slightly inferior to one of the 10 types of vehicles above since the high similarity of the T72 variant increases the difficulty of recognition. However, the accuracy is still satisfactory and higher than the previous work, further demonstrating the outstanding performance of our method. Thus, the umbrella model will be promise for the classification tasks of the SAR images.

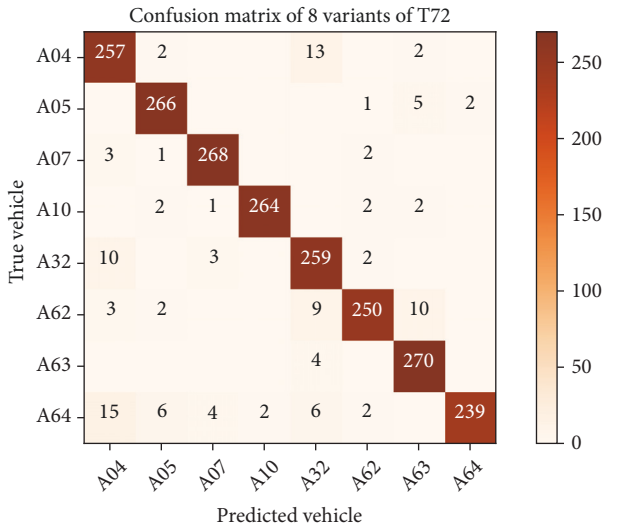

Figure 8: The confusion matrix of T72 8 variants. The accuracy of the test set is 96.35%.

4. Conclusions

Recently, the development of deep learning has been significantly advanced in the classification of optical images, in which many excellent CNN architectures emerge and achieve high performances. Compared to the optical recognition, the development of SAR-ART has been limited. Thus, it is highly desired to introduce advanced architectures into SAR-ART. However, since the SAR image presents fewer features than the optical one, these CNN architectures with high performance in the optical image easily cause overfitting for the SAR classification. Thus, in the work, based on the SAR characteristics, we constructed a novel CNN architecture (named umbrella) through minimizing its depth but extract enough features at different levels so as to achieve rapid and accurate detection of the SAR targets. Umbrella was applied to detect ten types of vehicles and eight classes of T72 variants from the MSTAR data set, where we also took into account the diverse positions of targets with the aid of the random translation manner. In all the cases under consideration, the umbrella model can achieve more than 99% accuracy for the classification of 10-class targets and higher than 96% accuracy for the 8 variants of the T72 tank. The performance of the umbrella model is higher than previous methods reported. The results clearly indicate that the architecture proposed by us will be potential for SAR-ATR in practice.

Data Availability

The authors would like to thank all the related agencies of these data providers; the data sources used in this paper can be obtained from https://www.sdms.afrl.af.mil/index.php?collection=mstar.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Xuemei Pu and Chuan Li supervised and designed the research, Yingjie Xie wrote the manuscript and performed the experiments, Yingjie Xie and Yijing Liu designed the methodology, Wenxin Dai developed related programs, Zhenxin Hu did data analysis, and Xuemei Pu reedited the manuscript and checked the experiment.

Acknowledgments

This research is supported by NSAF (Grant no. U1730127) and National Science Foundation of China (Grant no. 21573151).

References

[1] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[2] Z. Zhao, L. Jiao, J. Zhao, J. Gu, and J. Zhao, "Discriminant deep belief network for high-resolution SAR image classification," *Pattern Recognition*, vol. 61, pp. 686–701, 2017.

[3] X. Zhang, Z. Liu, S. Liu, D. Li, Y. Jia, and P. Huang, "Sparse coding of 2D-slice Zernike moments for SAR ATR," *International Journal of Remote Sensing*, vol. 38, no. 2, pp. 412–431, 2017.

[4] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1763–1767, 2017.

[5] D. E. Dudgeon and R. T. Lacoss, "An overview of automatic target recognition," *The Lincoln Laboratory Journal*, vol. 6, no. 1, pp. 3–10, 1993.

[6] Y. Cui, G. Zhou, J. Yang, and Y. Yamaguchi, "On the iterative censoring for target detection in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 641–645, 2011.

[7] L. M. Kaplan, "Analysis of multiplicative speckle models for template-based SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 4, pp. 1424–1432, 2001.

[8] K. Ikeuchi, M. D. Wheeler, T. Yamazaki, and T. Shakunaga, "Model-based SAR ATR system," vol. 2757 of *Proceedings of SPIE*, pp. 376–387, April 1996.

[9] H. Ma, J. Chan, T. K. Saha, and C. Ekanayake, "Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 2, pp. 468–478, 2013.

[10] G. J. Owirka, S. M. Verbout, and L. M. Novak, "Template-based SAR ATR performance using different image enhancement techniques," vol. 3721 of *Proceedings of SPIE*, pp. 302–319, April 1999.

[11] Y. Kuno, R. D. Juday, K. Ikeuchi, and T. Kanade, "Model-based vision by cooperative processing of evidence and hypotheses using configuration spaces," vol. 938 of *Proceedings of SPIE*, 444 pages, Orlando, Fl, USA, 1988.

[12] S. Singha, T. J. Bellerby, and O. Trieschmann, "Detection and classification of oil spill and look-alike spots from SAR imagery using an artificial neural network," in *Proceedings of the 2012 32nd IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2012*, pp. 5630–5633, Germany, July 2012.

[13] C. Yuan and D. P. Casasent, "MSTAR 10-Class classification and confuser and clutter rejection using SVRDM," in *Proceedings of the Defense and Security Symposium XVII*, pp. 624501–624513.

[14] J. Zhao, W. Guo, S. Cui, Z. Zhang, and W. Yu, "Convolutional neural network for SAR image classification at patch level," in *Proceedings of the 36th IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2016*, pp. 945–948, July 2016.

[15] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for SAR target classification with limited labeled training data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1091–1095, 2017.

[16] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sensing*, vol. 9, no. 9, p. 907, 2017.

[17] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[18] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 5, no. 2, pp. 157–166, 1994.

[19] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, pp. 770–778, 2015.

[21] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 1–9, IEEE, Boston, Mass, USA, June 2015.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 2818–2826, July 2016.

[23] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1800–1807, 2016.

[24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2016.

[25] G. Dong, N. Wang, and G. Kuang, "Sparse representation of monogenic signal: with application to target recognition in SAR images," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 952–956, 2014.

[26] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[27] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning (ICML '15)*, pp. 448–456, July 2015.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[29] S. Wagner, K. Barth, and S. Bruggenwirth, "A deep learning SAR ATR system using regularization and prioritized classes," in *Proceedings of the 2017 IEEE Radar Conference (RadarConf17)*, pp. 0772–0777, Seattle, WA, USA, May 2017.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[31] "Keras: The Python Deep Learning Library," 2018, https://keras.io/.

[32] K. P. Murphy, "Machine learning: a probabilistic perspective," *MIT*, 2012.

[33] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: a tutorial," in *Proceedings of the Algorithms for Synthetic Aperture Radar Imagery III*, vol. 2527, pp. 228–242, April 1996.

[34] K. Du, Y. Deng, R. Wang, T. Zhao, and N. Li, "SAR ATR based on displacement- and rotation-insensitive CNN," *Remote Sensing Letters*, vol. 7, no. 9, pp. 895–904, 2016.