Improved SAR feature fusion with convolutional neural networks and moment methods

Chunqian He [1], Dongsheng Li [2] and Yang Gao [3]

[1] National University of Defense Technology, He Fei, 230000, China; chunqianhe@nudt.edu.cn.

[2] National University of Defense Technology, He Fei, 230000, China; Lidsh@21cn.com.

[3] Information and Navigation College, Air Force Engineering University, Xi'an 710077, China; gao_yang_mail@163.com.

Corresponding author: Chunqian He (chunqianhe@nudt.edu.cn)

Key Points:

- Q_sigmoid function is superior to other methods for contrast enhancement of SAR images
- We propose an apporoch fusing moment features with convolution neural network features to improve the recognition rate of SAR images
- The recognition rate of our method is 98.63%, which is 2.77% higher than convolution neural networks and 6.32% higher than moment methods

Abstract

Synthetic aperture radar (SAR) is an important means of obtaining battlefield information. Correct identification of SAR images is essential. Hence, we propose a new method of SAR image recognition based on multi-feature fusion. Convolutional neural networks (CNNs) are based on local pixels and fuse this information at deep layers to obtain deep features. The moment method focuses on the entire image and obtains global moment features of different orders. We fuse the two feature types and choose a suitable feature fusion. First, we propose a Q_sigmoid function to enhance SAR image contrast, and then we separate the target area and remove noise interference. Next, we design a convolutional neural network to obtain the deep features of the target from the separated image. Then, we use four moment methods to produce the moment features of the image and sort them according to recognition rate. Finally, the above features are fused and sorted according to recognition rate to find the optimal combination. The recognition rate of our optimal fusion is 98.63%, which is 2.77% higher than is obtained by CNNs and 6.32% higher than moment methods.

1 Introduction

Synthetic aperture radar (SAR) is an active remote sensor with its own illumination; thus, it does not depend on light levels and can be used in all weather and in both day and night conditions (Brown, 1967). SAR images differ from optical images in several ways: (1) target size does not change with distance between sensor and target; (2) scene information is determined by amplitude of radar reflection and scattering; and (3) the image is highly sensitive to changes in target shape and attitude (El-Darymli et al., 2016). These characteristics make SAR a reliable instrument for collecting information, giving it important applications in both military and civilian fields (Cohen et al., 2016; Erten et al., 2016; Ugur et al., 2012). Many studies focus on SAR automatic target recognition (ATR) (He et al, 2014; Srinivas et al., 2014; Li et al., 2017), which can be divided mainly into feature-based methods and model-based methods (El-Darymli et al., 2016). Model-based methods focus on model design, but it is difficult to build a general target model for the multi-classification task. This limits the recognition rate of SAR images. Feature-based methods focus on extraction of target features; highly descriptive features can generate higher recognition rates. Thus, to enhance feature description, we explore the SAR image recognition method based on feature fusion.

Some classic feature-based methods, such as linear discriminant analysis (LDA) (Lu et al., 2003), principal component analysis (PCA) (Pei et al., 2016), and independent component analysis (ICA) (Wang et al., 2008), are sensitive to speckle noise and do not have rotation invariance, making these methods unsuitable for SAR image recognition. To overcome these problems, moment-based descriptors are applied to SAR image recognition. Raeisi (2018) extracted Zernike moment features of SAR images to distinguish oil from similar substances and achieved good results. Gishkori (2019) combined pseudo-Zernike moments with sparse representations for SAR target recognition. Pouya (2017) used Radial-Chebyshev moments on SAR target classification and achieved a 98.69% recognition rate for three target types. These successful applications indicate that moment features are suitable for SAR image recognition. Zernike moment (ZM) (Khotanzad & Hong, 1990), pseudo-Zernike moment (PZM) (Haddadnia & Faez, 2003), radial harmonic Fourier moment (RHFM) (Singh & Ranade, 2013), and Chebyshev-Fourier moment (CFM) (Zhu et al., 2016) have scale, translation, and rotation invariance in polar coordinates, making target recognition robust. Additionally, these four moment features yield feature extraction with low

computational cost and better performance than traditional features in SAR image recognition. Therefore, we implement these four moment methods for feature extraction. In addition to the moment feature, we implement convolutional neural networks (CNNs) in feature fusion.

CNNs have achieved impressive results in image recognition (Krizhevsky & Sutskever, 2012), Compared with traditional methods, a CNN uses different kernels to convolve images to obtain feature maps, and the feature maps go through stacked layers to obtain deep features (Szegedy, 2015). In contrast to a pre-designed manual feature, the deep feature shows better performance in optical image classification, due to the fact that the convolutional neural network is adaptive and does not require preset convolution kernels and parameters. This simplifies feature extraction without manual intervention and adjustment. Owing to the superiority of deep neural networks, several important studies on design of the deep learning architecture in SAR ATR have obtained excellent recognition rates (Chen et al., 2016; Ding et al., 2016; Kang et al., 2017). However, the deep feature itself does not have rotation, translation, or scale invariance. This invariance often requires corresponding transformation on the image dataset. For this reason, we propose integration of the moment feature to increase the invariance of the CNN feature.

By fusing the moment feature and the CNN feature of SAR images, we find the optimal fusion and training classifier. First, we preprocess the SAR image to separate the target area. Then, we extract 100-dimensional moment features from the processed image. The four types of moment feature are PZM, RHFM, ZM, and CFM. We simultaneously build a new CNN to extract 128-dimensional deep features of the image. The four moment features and deep features are then combined in different orders and sorted by their recognition rate. Finally, we select the feature combination with the highest recognition rate and evaluate its performance, while simultaneously comparing this method with other methods.

This article proceeds as follows. Section 2 introduces the dataset and the main steps of our experiment. Section 3 presents our experimental results, comparing different feature combinations and determining the optimal feature fusion. Section 4 evaluates our method's performance and compares it with other SAR ATR methods. Finally, Section 5 sets forth our conclusions.

2 Materials and Methods

Figure 1 shows a flowchart of our method. The input image goes through four steps: image preprocessing, feature extraction, several types of feature fusion and selection, and classification. In the following subsections, we preview the experimental dataset and the four steps.

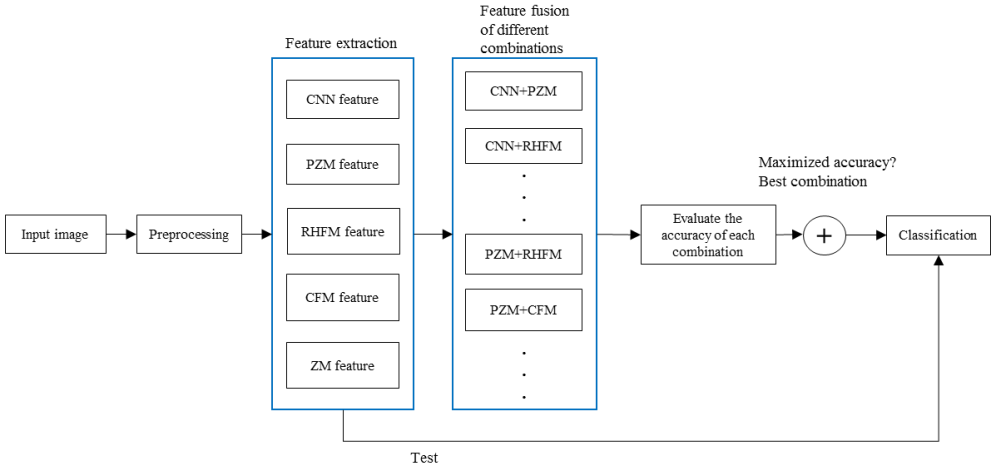


Figure 1. Flowchart of proposed method—Convolutional neural network (CNN), pseudo-Zernike moment (PZM), Zernike moment (ZM), radial harmonic Fourier moment (RHFM), Chebyshev-Fourier moment (CFM), "+" means find the optimal fusion.

2.1 MSTAR dataset

The MSTAR dataset is an SAR image set collected by high-precision beam-synthetic aperture radar, including a variety of Soviet military vehicles. The dataset is released by the US Defense Advanced Research Projects Agency (DARPA) and can be divided into standard operating condition (SOC) and extended operating condition (EOC). In this experiment, images under SOC are used, including 10 types of ground vehicle. Their collection pitch angles are 17° and 15°. Details of the dataset (target type, number of samples, collection pitch angle, picture size) are shown in Table 1. There are 2747 images at the 17° pitch angle and 2425 images at the 15° pitch angle, for a total of 5172 images.

Table 1.

MSTAR dataset for samples at different depression angles

Vehicle class	17° Depression angle	15° Depression angle	Image size
BMP2	233	195	128@128
BTR70	233	196	128@128
T72	232	196	128@128
BTR60	256	195	128@128
2S1	299	274	158@158
BRDM2	298	274	128@129
D7	299	274	177@178
T62	299	273	172@173
ZIL131	299	274	192@193
ZSU-23-4	299	274	158@158

2.2 Preprocessing

SAR images show serious interference from noise in the dataset. Each image includes the target, background noise, and target shadow area. Preprocessing is implemented to remove noise and shadow, keeping only the target area, called the RoI (region of interest). The characteristics of the SAR image indicate that the RoI has high gray levels compared with shadow and noise area. This allows division of image preprocessing into the following steps: contrast enhancement, mean filtering, morphological operation, threshold processing, and finally, center-cropping. The processing steps are shown in Figure 2.

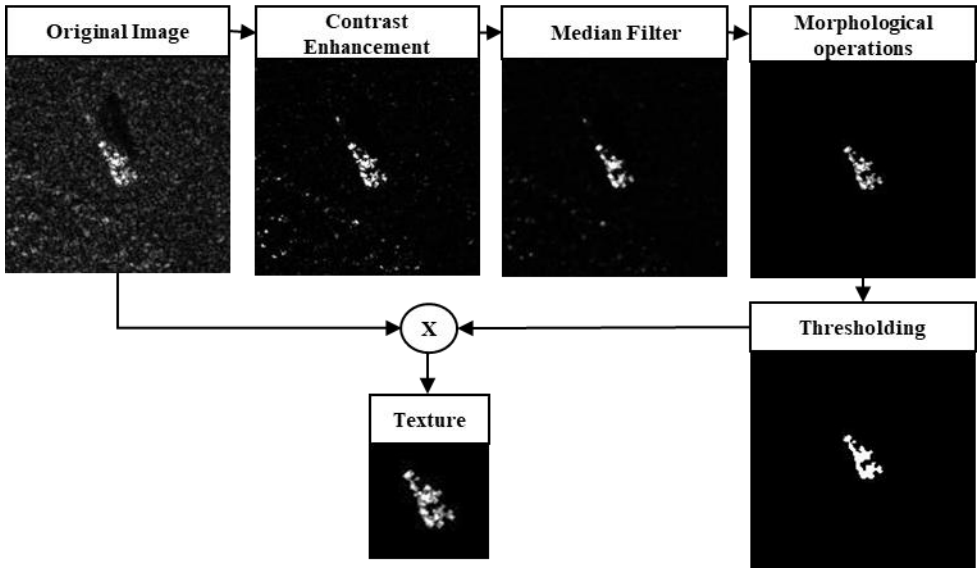


Figure 2. Segmentation process of a vehicle

Existing contrast enhancement methods include the histogram method and the grayscale transformation method. Although these methods improve image contrast, noise is also transformed. Thus, we propose a new nonlinear grayscale transformation approach in which the mapping function is a variant of the sigmoid function. We call this the Q_sigmoid function:

$$f(x) = \frac{1}{1 + e^{-k \times (x - 0.5)}} \tag{1}$$

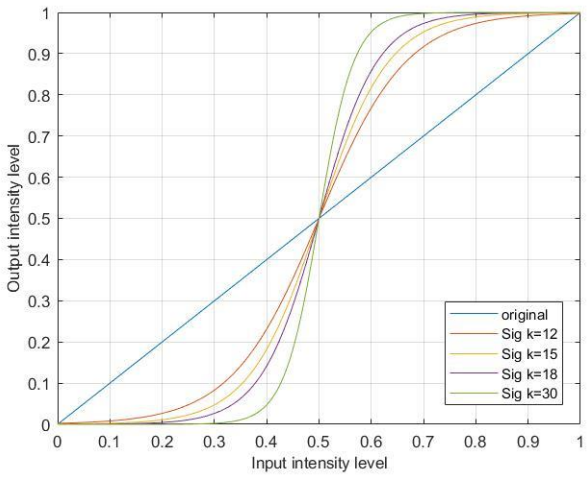
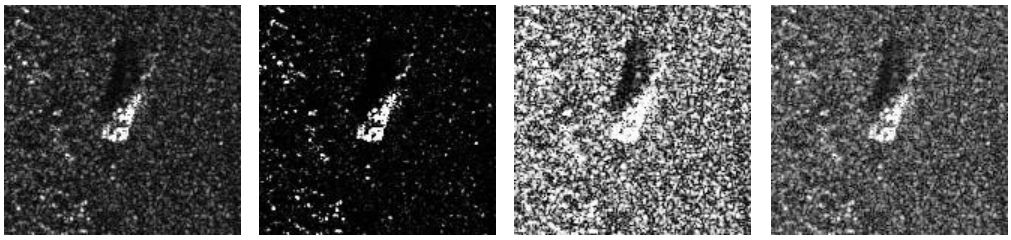


Figure 3. Q_sigmoid function curves with differing k

Figure 3 shows a set of curves when Q_sigmoid function parameter k takes different values. The horizontal axis represents the gray level of the original image, and the vertical axis is the gray level after mapping. It is possible that high gray level will increase through mapping, and low gray level will decrease; thus, image contrast could be enhanced. Our procedure sets the parameter $k = 12$. Figure 4 shows the results of SAR image contrast enhancement using different methods. Histogram equalization (c) brightens the noise, log transformation (d) fails to separate the background, while our method (b) highlights the target area (RoI) clearly.



a) Original image b) Q_sigmoid transformation c) Histogram equalization d) Log transformation

Figure 4. Contrast enhancement via different methods

After contrast enhancement, most noise is removed through filtering, but some information in the RoI area is also removed; it is necessary to correct this later. A morphological operation—bwmorph (a command in the MATLAB toolbox)—is used for hole filling. Thresholding is then performed to obtain the binary map representing the RoI area. Due to shadow and noise, gray levels become very low after contrast enhancement; a low threshold should be accepted to retain the greatest possible RoI. We find that the proper threshold is in the range [0.5, 2.5]; this should be fine-tuned according to SAR working conditions. Finally, the binarized logical map of the RoI area is merged with the original image to obtain the segmented target image. It is then center-cropped to acquire the target image with standard size 64 @ 64.

2.3 Feature extractuon based on CNN

In recent years, new networks have been proposed (Gu et al., 2018). Here, we refer to the He Kaiming (2016) Resnet network structure to create a CNN for feature extraction. This network features faster training and good image recognition performance. The network structure is shown in Table 2. a @ b represents the size of the kernel or graph. The parameters in 'Output' are channel_out and feature map size; channel_out represents the number of convolution kernels or the number of feature maps. The stride defaults to 1 if we do not instruct it. Convolution kernel and pooling kernel default size is 3 @ 3.

Image input convolution network size is (1, 64 @ 64). The Pre_layer uses 5 @ 5 convolution kernels to obtain 32 feature maps. The size of the feature map is further reduced through maxpooling. Resblock is continuously stacked in Layers 1, 2, and 3 to improve network depth, which can improve recognition performance. Finally, the convolutional neural network converts 128 feature maps into a vector in the feature layer through average pooling. The 128-dimensional vector is the feature extracted by CNN.

Table 2.

CNN structure

Layer name	Structure	Output
Pre-layer	Conv (32,5@5, stride=2)	(32,32@32)
	Max pool (2@2, stride=2)	(32,16@16)
Layer1	Resblock1	(32,16@16)
	Resblock2	(32,16@16)
Layer2	Resblock1 (stride=2)	(64,8@8)
	Resblock2	(64,8@8)

Layer3	Resblock1 (stride=2)	(128,4@4)
	Resblock2	(128,4@4)
Feature layer	Average pool (2x2)	(128@1)

Resblock refers to the shortcut structure of He. As shown in Figure 5, channel_out represents the number of feature maps after convolution; the default stride is 1. The stride of the first convolution layer in Figure 5 needs to be adjusted if we note in table 2 “structure” column. A shortcut structure can speed up the network without loss of depth.

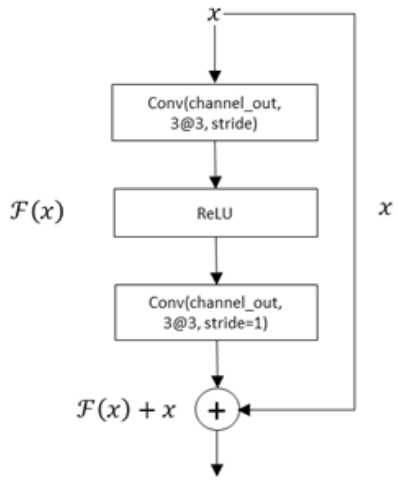


Figure 5. Structure of a Resblock

During the CNN training process, we divide the dataset into three parts: 60% of images are used as the training set, 20% as the verification set, and the remaining 20% as the test set. Training continues until training loss is close to 0. Using CNN features to classify SAR images yields a recognition rate as high as 95.86%, and a total of 4963 images are correctly identified.

2.4 Feature extraction based on moment methods

For images, pixel coordinates can be regarded as a two-dimensional random variable (X, Y). A grayscale image can then be represented by a two-dimensional gray density function, so moments can be used to describe the characteristics of gray images (Teague, 1980). We use ZM, PZM, CFM, and RHFM to extract image features. The generalized moment function is as follows in Cartesian coordinates:

$$M_{p,q} = \int\int P_p(x) P_q f(x,y) dxdy\,, \tag{2}$$

where $P_p(x)$ and $P_q(y)$ are the polynomial basis, $f(x,y)$ is the digital image, p is order, and q is repetition. The four moment methods have different basis functions. Due to the fact that radial basis functions are completely orthogonal within the unit circle, we use a moment function in polar coordinates as

$$M_{p,q} = \int_0^{2\pi}\int_0^1 R_{p,q}(\rho) e^{-jq\theta} f(\rho,\theta)\rho d\rho d\theta\,. \tag{3}$$

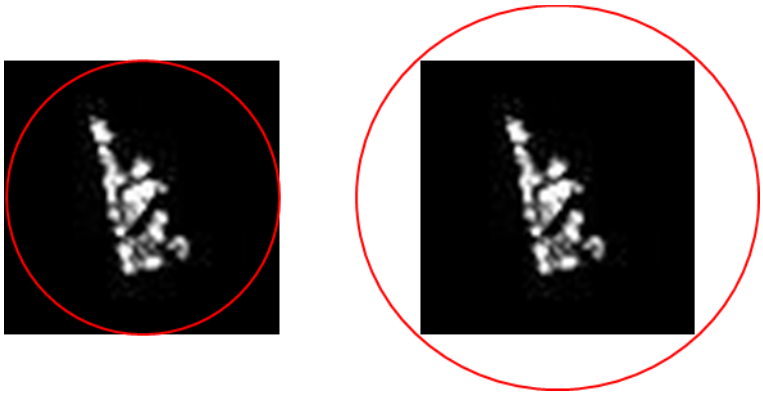


Figure 6. Two mapping relations of an image in a unit circle

To make the extracted moment features invariant, it is necessary to map the image into the unit circle before calculating the moment. There are two ways to do this, as shown in Figure 6: an inscribed circle and a circumscribed circle (Cahndan & Sukhjeet, 2013). Generally, we would consider the circumscribed circle so that the moment feature will contain all image information. However, in this study, the inscribed circle is better, as the image outside the circle contains no useful information. For $N \times N$ discrete digital images, we map pixels in the inscribed circle to the unit circle using the following mapping function:

$$\begin{cases} \rho = \dfrac{\sqrt{(2x - N + 1)^2 + (2y - N + 1)^2}}{N} \\ \theta = \tan^{-1}\left(\dfrac{N - 1 - 2y}{2x - N + 1}\right) \end{cases}. \tag{4}$$

For discrete digital images, the moment calculation formula is written as:

$$M_{p,q} = \frac{\pi}{\lambda_N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(\rho, \theta) R_{p,q}(\rho) e^{-jq\theta}, \tag{5}$$

where λ_N is the normalization factor, which equals the number of image pixels mapping in the unit circle, and the corresponding continuous case is π.

2.5 Classifier

Pouya (2017) used SVM (support vector machine) as the classifier when applying moment features for SAR image recognition. SVM is suitable for clustering and classification problems because it is robust and sparse. However, in the present study, we introduce multilayer perceptron (MLP) (Dhlamini & Marwala, 2004) as the classifier, due to its high accuracy and its applicability in neural networks. MLP simulates the human brain and consists of multiple interconnected neural nodes. As shown in Figure 7, connections between nodes at different layers have different weights. Each weight represents the influence of one node on another node. The output of the previous layer is stimulated or suppressed with respect to the next layer at different levels. The number of output layer nodes is equal to the number of categories, and the number of input layer nodes is equal to the dimensions of the feature vector.

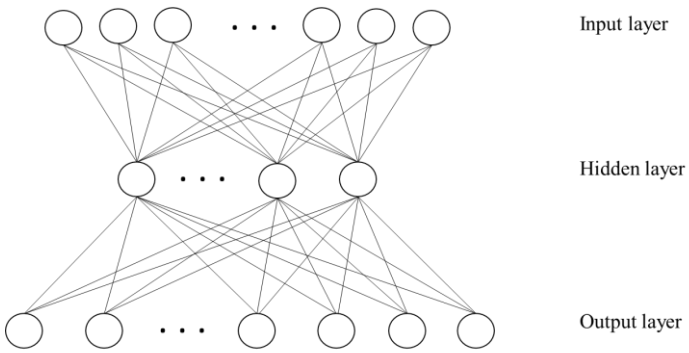


Figure 7. Multilayer Perceptron classifier structure

The weight parameters of the MLP classifier need to be learned from the data, so network training requires a larger number of samples. We compare recognition performance using SVM and MLP as classifiers. The MLP has three layers, 100 nodes in the input layer (equal to the length of the moment feature), 10 nodes in the output layer (equal to the number of categories), and 30 nodes in the hidden layer. As shown in Table 3, for the same moment feature, the recognition rate from using the MLP classifier is improved by 2%–5% compared with SVM. For this reason, we apply a three-layer MLP as the classifier in subsequent feature fusion experiments.

Table 3.
Accuracy (%) of each method in the dataset using different classifiers

Method	Feature dimension	SVM classifier accuracy (%)	MLP classifier accuracy (%)
Pseudo-Zernike moments	100	88.51	92.56
Radial harmonic Fourier moments	100	86.17	89.02
Chebyshev-Fourier moments	100	83.23	88.44
Zernike moments	100	84.87	86.97
Convolutional neural network	128	\	95.86

3 Experiment and discussion

In this section, we fuse different moment features (*fusion* refers to the splicing of feature vectors) and train the MLP for classification. We then sort these features according to recognition rate to find the optimal fusion. Then, we combine the CNN feature with moment features in different orders, and train the MLP classification to get a recognition rate. We then select the combination with the highest recognition rate.

We begin with the fusion of moment features. There are four kinds of moment feature. Moment feature PZM is far superior to ZM and is a variant of ZM; therefore, ZM is not considered when performing fusion. After fusion, the dimensions of the feature increase, which means that the number of hidden layer nodes of the MLP classifier needs to be adjusted. When the input feature dimensions are 300, 200, and 100, the corresponding hidden layer nodes are 60, 50, and 30. The learning rate is e-3 at the start of training; after 10 epochs it is adjusted to e-4 and training continues until the loss approaches 0. This procedure reduces training time without reducing training results.

Table 4.

Accuracy of fusion of different moment features—pseudo-Zernike moment (PZM), Zernike moment (ZM), radial harmonic Fourier moment (RHFM), Chebyshev-Fourier moment (CFM), "+" means feature fusion.

Rank	Method	Number of features	Accuracy (%)
1	PZM+CFM	100+100	93.31
2	PZM	100	92.56
3	PZM+RHFM	100+100	90.18
4	RHFM	100	89.02
5	RHFM+CFM	100+100	88.90
6	CFM	100	88.44
7	ZM	100	86.97

Moment feature fusion and its recognition rates are shown in Table 4. The performance of the single moment feature is PZM > RHFM > CFM for the SAR image recognition task. For the combination of moment features, PZM+RHFM does not improve the recognition rate and performs worse than the single PZM feature. Fusion feature RHFM + CFM does not perform better than the single RHFM feature, so the RHFM and PZM features are not suitable for fusion. Similarly, RHFM and CFM are not suitable for fusion Thus, we consider the fusion of PZM and CFM. Our experiment shows that the fusion feature PZM + CFM is superior to the PZM feature and shows the highest recognition rate. As a result, for SAR image recognition, the best moment feature is the PZM feature, and the optimal feature fusion is PZM + CFM.

Table 5. Accuracy of fusion of CNN features and moment features—Convolutional neural network(CNN), pseudo-Zernike moment(PZM),Zernike moment(ZM),radial harmonic Fourier moment(RHFM), Chebyshev-Fourier moment(CFM), "+" means feature fusion.

Rank	Method	Number of features	Accuracy (%)
9	CNN	128	95.86
6	CNN+PZM	128+100	98.01
5	CNN+RHFM	128+100	98.11
8	CNN+ZM	128+100	96.79
7	CNN+CFM	128+100	97.29
2	CNN+PZM+RHFM	128+100+100	98.53
1	CNN+PZM+CFM	128+100+100	98.63
3	CNN+RHFM+CFM	128+100+100	98.36
4	CNN+PZM+RHFM+CFM	128+100+100+100	98.34

Compared with the moment feature, the deep feature extracted via CNN has a very high recognition rate. The attributes of these two kinds of feature differ. The moment feature is a global feature, and the CNN feature is a semantic feature which obtains from the local field through stacking layers. Table 5 shows the different combinations of the two features. It is possible that the fusion of the moment feature and CNN feature improves the recognition rate. The reason for this is that the moment feature complements the deep feature, and its invariance improves deep

feature distinction. In Section 2, we discussed the fusion of moment features and concluded that the single moment feature PZM is best, and that the fusion moment feature PZM + CFM performs best. There is little difference when it is fused with the CNN feature. Comparing with single CNN feature, when one moment feature is fused, CNN + RHFM performs best and the recognition rate increases by 2.25%; With two moment features fused, CNN + PZM + CFM performs best, and the recognition rate increases by 2.77%. This result is consistent with the moment feature fusion result discussed in Section 2; that is, PZM + CFM is the best fusion moment feature. The final recognition rate is 98.63% for the CNN feature fused with moment features.

In addition, with respect to MLP training time requirements, increasing the feature dimension does not greatly increase the training time of the classifier. A well-trained network can complete recognition of an input image almost in real time. Further, this experiment was conducted on a laptop, with CPU i5-8300H, NVIDIA GeForce GTX 1050Ti GPU, 8G memory. CNN + PZM + CFM classifier training time is 30 minutes and CNN classifier training time is 40 minutes. In terms of recognition rate improvement, such time requirements are justified.

4 Performance evaluation and comparison

We compare our method with recent SAR image recognition research. Table 6 shows recognition rates for the following methods: classic template matching methods (Deng et al., 2017); NFM method (Cui et al., 2015); methods based on neural networks, i.e., original autoencoder and Euclidean distance restricted autoencoder (Deng et al., 2017); sparse representation method, i.e., JMSDR (Yu et al., 2019); moment feature fusion method, i.e., moment methods based on Fisher score (Pouya et al, 2020); and deep CNN method, i.e., 4-VDCNNs (Pei et al., 2018). Our method shows the highest recognition rate. Although the neural network itself has strong recognition capabilities, the moment feature's rotation and translation invariance further enhance discrimination of the CNN feature. Thus, the recognition rate is improved.

Table 6.

Recognition rate for each method

Method	Accuracy (%)
Euclidean distance restricted autoencoder	91.29
Original Auto Encoder	87.04
NMF	95.16
Monogenic signal	92.79
Template matching	87.58
Moment methods based on Fisher score	92.48
4-VDCNNs	98.52
JMSDR	93.18
Proposed method	98.63

In addition to comparing the accuracy rate, we also evaluate the stability of our method. The important indicators of this evaluation are confusion matrix and true-positive rate (TPR) /

sensitivity, true-negative rate (TNR) / specificity, false-negative rate (FNR) / miss rate, and accuracy (ACC), defined as follows:

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{6}$$

$$\text{Specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{7}$$

$$\text{MissRate} = \text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}}, \tag{8}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}. \tag{9}$$

TP, FN, TN, and FP represent true-positive, false-negative, true-negative, and false-positive results, respectively.

The confusion matrix clearly indicates recognition of each type of target and is related to the four indicators described above. The diagonal elements of Figure 8 represent the number of correct recognitions, or TP. Each column of the matrix represents the predicted class, and each row represents the actual class. In addition to the diagonal elements (TP), in the column direction, the values represent the number of images that the algorithm misjudges other classes as a predict class (FP). In the row direction, the values represent that the actual class is wrongly predicted as other predicted classes (FN). The figure shows the relationship between the confusion matrix and TP, TN, FP, and FN. The red line divides the confusion matrix into four parts. The sum of values in each part of the area is the TP, TN, FP, and FN of one class.

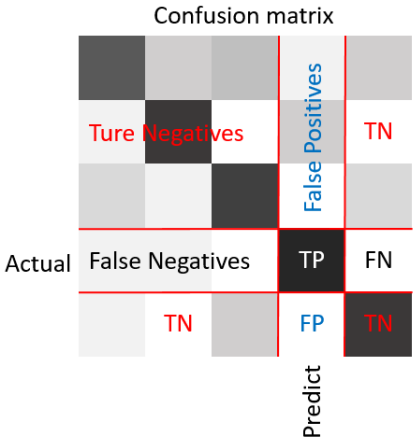


Figure 8. Confusion matrix and TP, TN, FN, FP—true positive (TP), ture negative(TN),false negative(FN), false positive(FP).

We create a confusion matrix of CNN feature recognition, moment feature (PZM + CFM) recognition, and CNN + moment feature (CNN + PZM + CFM) recognition. As shown in Figures 9–11, CNN + PZM + CFM recognition performs best; out of a total of 5172 images, 5101 are correctly classified. CNN recognizes 4958 images, and PZM + CFM recognizes only 4870 images. In addition, the CNN + moment feature method shows the best performance on each type of target classification; only one false occurs on D7, and the highest false recognition occurs on BTR60 out of 14 false recognitions. Compared with CNN + PZM + CFM, the CNN feature method recognizes three types of targets well, and the PZM + CFM feature fusion has lower recognition rate on each category.

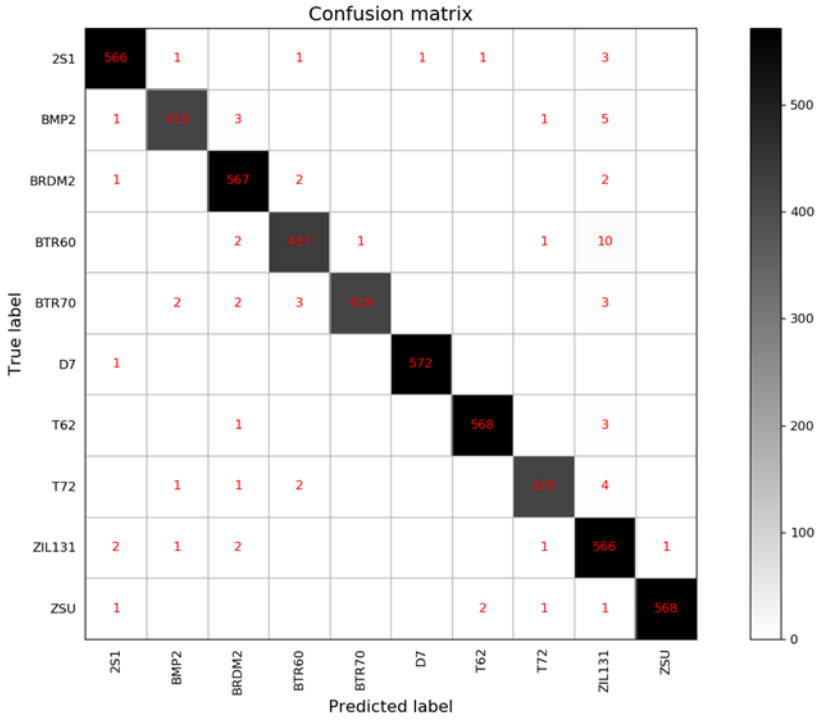


Figure 9. Confusion matrix of CNN + PZM + CFM classification

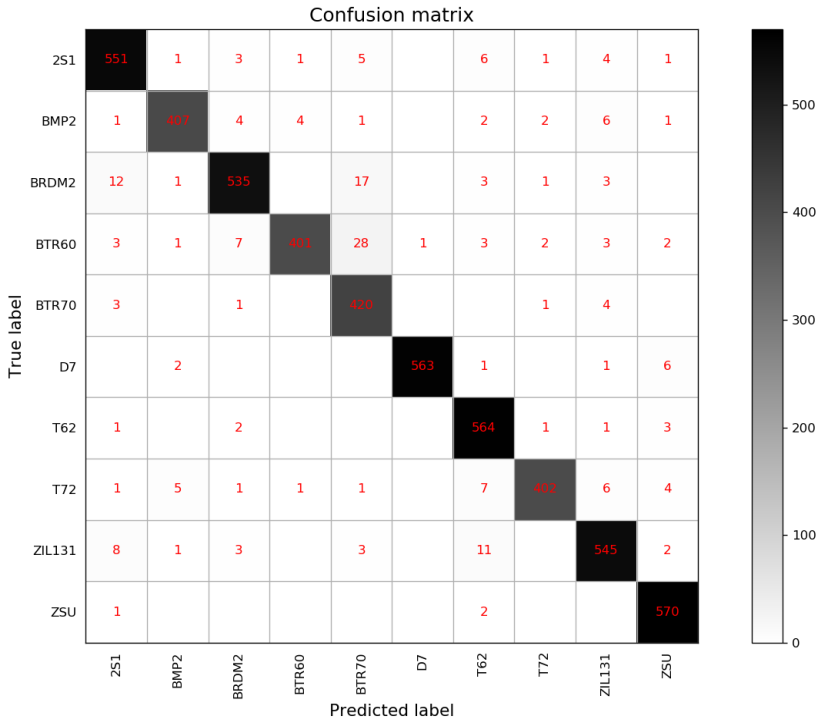


Figure https://doi.org/10. Confusion matrix of CNN classification

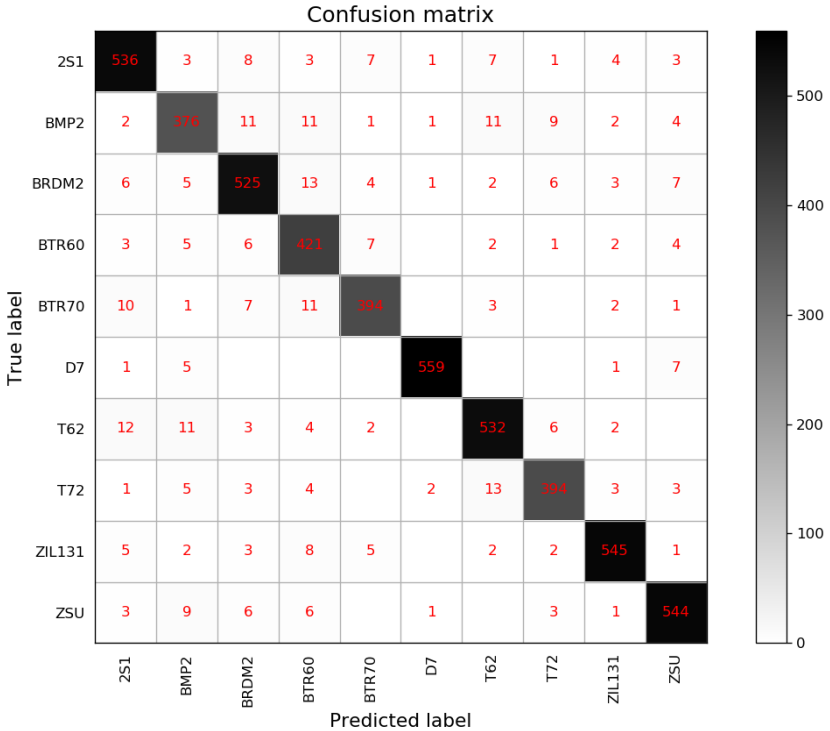


Figure 11. Confusion matrix of PZM + CFM classification

Table 7.

Evaluation of CNN + PZM + CFM using accuracy, TPR, TNP, and FNR—true positive rate(TPR)， true negative rate(TNR), false negative rate(FNR).

Target category	2S1	BMP2	BRDM2	BTR60	BTR70	D7	T62	T72	ZIL131	ZSU
Accuracy	99.75	99.71	99.69	99.57	99.79	99.96	99.86	99.77	99.27	99.88
TPR	98.78	97.66	99.13	96.90	97.97	99.82	99.30	98.13	98.78	99.13
TNR	99.87	99.89	99.76	99.83	99.98	99.98	99.94	99.92	99.33	99.98
FNR	1.22	2.34	0.87	3.10	2.33	0.17	0.70	1.82	1.22	0.87

Table 7 shows the accuracy, TPR, TNR, and FNR of each target using the best fusion method. Accuracy is above 99% for each type of target, minimum TPR is 96.90%, and TNR is above 99%. The maximum FNR of 3.1% indicates that the method has good recognition performance for each category. We compare these four indicators on the three methods. For clarity, we take only the maximum value, minimum value, and average value for comparison. As shown in Table 8, they all have better performance on accuracy, but CNN + PZM + CFM on TPR is better than the other two methods and FNR is much smaller. Thus, the CNN + PZM + CFM method not only achieves a 98.63% recognition rate, it also has good performance in each category. The classifier's ability is greatly improved using this fused feature.

Table 8.

Evaluation of three methods using accuracy, TPR, TNR, and FNR—true positive rate(TPR)， true negative rate(TNR), false negative rate(FNR).

Method	Average				Maximum				Minimum			
	ACC	TPR	TNR	FNR	ACC	TPR	TNR	FNR	ACC	TPR	TNR	FNR
CNN	99.18	95.70	99.54	4.29	99.79	99.48	99.98	11.09	98.76	88.91	98.84	0.52
PZM+CFM	98.66	93.10	99.26	6.90	99.61	97.56	99.87	12.15	98.11	87.85	98.73	2.44
CNN+PZM+CFM	99.73	98.56	99.85	1.46	99.96	99.82	99.98	3.1	99.27	96.90	99.33	0.17

5 Conclusions

This paper proposes a novel approach toward SAR image recognition. Our method is based on CNN and moment feature fusion, which can improve the recognition rate by enhancing feature description. First, we propose a Q_sigmoid function to improve SAR image contrast. We then compare the SVM and MLP classifiers and show that MLP is more suitable for this task. Our feature fusion experiment shows that the optimal fusion feature is CNN + PZM + CFM. By fusing the deep feature and moment feature, a high recognition rate of 98.63% is achieved. We compare our method with several existing methods, and the result shows our method to be superior. Finally, our method shows high stability and adaptability. In future work, we will explore the fusion of CNN features and moment features, not only stitching together the two features, but merging with trade-offs, while determining the appropriate fusion weight of the features.

Author Contributions: Chunqian He: Conceptualization of this study, Methodology, Software coding and manipulation, Validation Methodology, Writing—Original draft preparation. Yang

Gao: Review of the methodology and validation process. Dongsheng Li: Supervisor. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank National University of Defense Technology for the support with this research.

Data availability statement： Data is available through Pouya et al. 2020 and Pei et al. 2018.

Conflicts of Interest: The authors declare no conflict of interest.

References

Brown, W. M.(1967). Synthetic aperture radar. *IEEE transactions on aerospace and electronic systems*.3, 217–229.

El-Darymli, K., Gill, E.W., McGuire, P., Power, D.& Moloney, C(2016). Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review. *IEEE Access.* 4, 6014-6058. https://doi.org/10.1109/ACCESS.2016.2611492

Cohen, J., Riihimaki, H., Pulliainen, J., Lemmetyinen, J.&Heilimo, J(2016). Implications of boreal forest stand characteristics for X-band SAR flood mapping accuracy. *Remote Sensing of Environment.* 186, 47–63. https://doi.org/10.1016/j.rse.2016.08.016

Erten, E., Lopez-Sanchez, J.M., Yuzugullu, O. et al.(2016). Retrieval of agricultural crop height from space: a comparison of SAR techniques. *Remote Sensing of Environment.* 187, 130–144. https://doi.org/10.1016/j.rse.2016.https://doi.org/10.007

Ugur, S., Arikan, O(2012). SAR image reconstruction and autofocus by compressed sensing. *Digital Signal Process*ing. 22(6), 923–932. https://doi.org/10.1016/j.dsp.2012.07.011

He, Y., He, S.Y., Zhang, Y.H., Wen, G.J., Yu, D.F., Zhu, G.Q(2014). A forward approach to establish parametric scattering center models for known complex radar targets applied to SAR ATR. *IEEE Transactions on Antennas and Propagation.* **2014**, 62(12), 6192–6205. https://doi.org/10.1109/TAP.2014.2360700

Srinivas, U., Monga, V., Raj, R.G(2014).SAR automatic target recognition using discriminative graphical models. *IEEE Transactions on Aerospace and Electronic Systems*, 50(1), 591–606. https://doi.org/10.1109/TAES.2013.120340

Li, C., Bao, W., Xu, L., Zhang, H (2017). Clustered multitask learning for automatic radar target recognition. *Sensors*. 17(10), 2218. https://doi.org/10.3390/s17102218

Lu, J., Plataniotis, J.K.N., Venet sanopoulos, A.N(2003). Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks*. 14(1), 195–200. https://doi.org/10.1109/TNN.2002.806647

Pei, J., Huang, Y., Huo, W., Wu, J., Yang, J., Yang, H(2016). SAR imagery feature extraction using 2DPCA-based two-dimensional neighborhood virtual points discriminant embedding. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), 2206–2214. https://doi.org/10.1109/JSTARS.2016.2555938

Wang, H., Pi, Y., Liu, G., Chen, H(2008). Applications of ICA for the enhancement and classification of polarimetric SAR images. *International Journal of Remote Sensing*. 29(6), 1649–1663. https://doi.org/10.1080/01431160701395211

Raeisi, A., Akbarizadeh, G., Mahmoudi(2018), A. Combined Method of an Efficient Cuckoo Search Algorithm and Nonnegative Matrix Factorization of Different Zernike Moment Features for Discrimination Between Oil Spills and Lookalikes in SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* 11(11), 4193-4205. https://doi.org/10.1109/JSTARS.2018.2841503

Gishkori, S., Mulgrew, B(2019). Pseudo-Zernike Moments Based Sparse Representations for SAR Image Classification. *IEEE Transactions on Aerospace and Electronic Systems.* 55(2), 1037-1044. https://doi.org/10.1109/TAES.2018.2856321

Pouya, B., Hasan, D., Sener, U(2017). Target recognition in SAR image using radial Chebyshev moments. *Signal Image Video Processing.* 11(6), 1033-1040. https://doi.org/10.1007/s11760-017-1054-2

Khotanzad, A., Hong, Y.H(1990). Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 12(5), 13–118. https://doi.org/10.1109/34.55109

Haddadnia, J., Faez, K., Ahmadi, M(2003). An efficient human face recognition system using pseudo Zernike moment invariant and radial basis function neural network. *International Journal of Pattern Recognition and Artificial Intelligence.* 17(1), 41–62. https://doi.org/10.1142/S0218001403002265

Singh, C., Ranade, S.K(2013). A high capacity image adaptive watermarking scheme with radial harmonic Fourier moments. *Digital Signal Processing*. 23(5), 1470–1482. https://doi.org/10.1016/j.dsp.2013.05.006

Zhu, H., Yang, Y., Gui, Z., Zhu, Y., Chen, Z(2016). Image analysis by generalized Chebyshev–Fourier and generalized pseudo Jacobi–Fourier moments. *Pattern Recognition.* 51, 1–11. https://doi.org/10.1016/j.patcog.2015.09.018

Krizhevsky, A., Sutskever, I., Hinton, G(2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 2, 1097-1105.

Szegedy, C., Liu, W., Jia, YQ., Sermanet, P. et al(2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 7.12, 1-9 https://doi.org/10.1109/cvpr.2015.7298594

Chen, S., Wang, H., Xu, F., Jin, Y.Q(2016). Target classification using the deep convolutional networks for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*. 54(8), 4806–4817. https://doi.org/10.1109/TGRS.2016.2551720

Ding, J., Chen, B., Liu, H.W., Huang, M.Y(2016). Convolutional Neural Network with Data Augmentation for SAR Target Recognition. *IEEE Geosciense and Remote Sensing Letters.* 13(3), 364-368. https://doi.org/10.1109/LGRS.2015.2513754

Kang, M., Ji, K., Leng, X., Xing, X., Zou, H(2017). Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder. *Sensors***,**17(1), 192. https://doi.org/10.3390/s17010192

Gu, J., Wang, Z., Kuen, J., et al.(2018). Recent advances in convolutional neural networks. *Pattern Recognition.* 77, 354-377. https://doi.org/10.1016/j.patcog.2017.https://doi.org/10.013

He, K., Zhang, X., Ren, S. et al. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. 770-778. https://doi.org/10.1109/CVPR.2016.90

Teague, M.R(1980). Image analysis via general theory of moments. Journal of the Optical Society of America. 70(8), 920-930. https://doi.org/10.1364/JOSA.70.000920

Cahndan, S., Sukhjeet K.R(2013). A High capacity image adaptive watermarking scheme with radial harmonic Fourier moments. *Digital Signal Processing*. 23(5),1470-1482. https://doi.org/10.1016/j.dsp.2013.05.006

Dhlamini, SM., Marwala T(2004). An application of SVM, RBF and MLP with ARD on bushings. *IEEE Conference on Cybernetics and Intelligent Systems*, 1253-1258.

Deng, S., Du, L., Li, C., Ding, J., Liu, HW(2017). SAR Automatic Target Recognition Based on Euclidean Distance Restricted Autoencoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* 10(7), 3323-3333. https://doi.org/10.1109/JSTARS.2017.2670083

Cui, Z., Cao, Z., Yang, J., Feng, J., Ren, H(2015). Target recognition in synthetic aperture radar images via non-negative matrix factorization. *IET Radar, Sonar and Navigation*. 9(9), 1376–1385. https://doi.org/10.1049/iet-rsn.2014.0407

Pouya, B., Masoud, M., Hasan, D., Sener, U(2020). Improved SAR target recognition by selecting moment methods based on Fisher score. *Signal Image Video Processing*. 14(1), 39-47. https://doi.org/10.1007/s11760-019-01521-5

Pei, J., Huang, Y., Huo, W., Zhang, Y., Yang, J., Yeo, T-S(2018). SAR Automatic Target Recognition Based on Multiview Deep Learning Framework. *IEEE Transactions on Geoscience and Remote Sensing*. 56(4), 2196–22https://doi.org/10. https://doi.org/10.1109/TGRS.2017.2776357

Yu, M., Quan, S., Kuang, G., Ni, S(2019). SAR Target Recognition via Joint Sparse and Dense Representation of Monogenic Signal. *Remote Sensing*. 11(22), 2676. https://doi.org/10.3390/rs11222676