(Received 20 June 2024, Accepted 22 August 2024, Scheduled 3 September 2024)

SAR Target Recognition Based on Multi-View Differential Feature Fusion Network Under Small Sample Conditions

Yuxin Ma[1], Benyuan Lv[1], Jianfei Ren[1], Yun Guo[2], Jiacheng Ni[1,*], and Ying Luo[1]

[1] *Information and Navigation School, Air Force Engineering University, Xi'an 710082, China*
[2] *College of Physics & Electronic Information, Luoyang Normal University, Luoyang 471023, China*

ABSTRACT: Deep learning network has the advantages of strong learning ability, strong adaptability, and good portability. Therefore, synthetic aperture radar (SAR) automatic target recognition (ATR) based on deep network is widely used in both military and civilian fields. However, due to the imaging conditions, radar angle, imaging distance, and other reasons, it is difficult to obtain efficient and usable SAR image datasets. SAR images' recognition under small sample conditions is still a challenging problem. In this paper, a SAR target recognition method based on multi-view differential feature fusion network is proposed to address this problem. Considering the correspondence between RCS and target features, the network extracts dissimilarities between features from SAR images of different angles of the same target and fuses them with the original features of one angle to form new features, which enriches the available training data. Experimental results on the Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset show that the proposed method has a higher target recognition rate than other deep network methods, as well as single angle input recognition methods.

1. INTRODUCTION

Synthetic aperture radar (SAR) imaging has high military and civilian values due to its all-weather, all-day, electromagnetic wave penetration, and light-independent characteristics. Target recognition technology based on SAR images is widely used in military operations, meteorological monitoring, geological exploration, urban management, and other fields [1, 2].

The current SAR automatic target recognition (ATR) is mainly divided into three methods: template-based, model-based, and deep learning-based ones. The template-based recognition method [3] is simpler, but requires the collection of a large number of measured samples, which is difficult to achieve. The model-based recognition method [4] does not need to collect a large number of measured samples, but the method is computationally intensive and difficult to meet real-time requirements. Compared with the first two methods, deep learning-based recognition algorithms can automatically learn features and achieve end-to-end training. Both recognition accuracy and real time performance are better than traditional methods. Therefore, recognition algorithms based on deep learning have become the current mainstream [5]. However, the acquisition of large scale SAR images requires huge financial and material resources, so SAR target recognition based on deep learning often faces small sample problem [6]. Therefore, it is of great theoretical and practical value to study how to achieve better SAR recognition under small sample conditions. At present, convolutional neural network [7], deep confidence network, recurrent neural network, etc. [8, 9] are applied in SAR target recognition, but there are still many problems; especially when the number of samples is scarce, the recognition performance of the algorithm needs to be further improved [10].

In recent years, a number of innovative approaches have been proposed with the aim of addressing the issue of insufficient training images in SAR target recognition. Chen et al. [11] augmented the training dataset by randomly cropping and flipping the images in the dataset, which enhanced the accuracy of SAR recognition to a certain extent. However, the data augmentation had limited effect when the samples were very small, and it did not fundamentally increase the features learned by the network. Consequently, this did not result in a significant increase in recognition accuracy, but it did lead to an increase in training time [12]. Bao et al. [13] proposed a Spectral Normalization-Generative Adversarial Network (SN-GAN)-based SAR image simulation method to generate SAR images with high similarity to real SAR images. However, the SN-GAN algorithm is unable to guarantee the quality of the generated images when the number of original SAR images is small, which may degrade the recognition accuracy of SAR. While all of the aforementioned methods improve the network's recognition ability of SAR images in the case of small samples to a certain extent, the performance improvement of the above methods is limited when the original samples are extremely scarce [14, 15]. In particular, the majority of current algorithms are single-input SAR target recognition networks. However, due to the unique imaging mechanism of SAR, the features of SAR images from different angles are markedly distinct. A single-input network is unable to leverage the feature information embedded in different angles, which results in suboptimal network recognition performance when the samples are limited. In recent years, some scholars have proposed that different angles of SAR images contain rich information. If these features' information can be

* Corresponding author: Jiacheng Ni (littlenjc@sina.com).

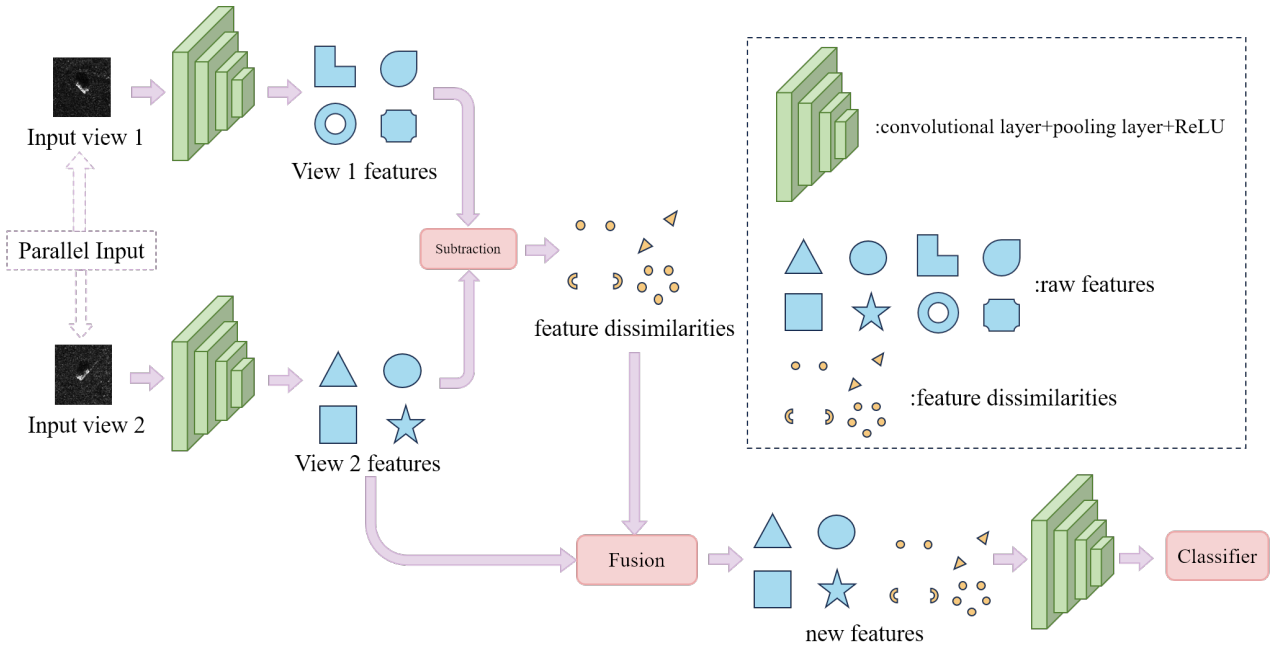

FIGURE 1. Basic architecture of the proposed Multi-view Differential Feature Fusion Network (MDFFN).

effectively extracted and utilised, it will lead to an improvement in the recognition performance. Pei et al. [16] proposed a Deep Convolutional Neural Network (DCNN) algorithm, which extracts the features of different angles by fusing the features of SAR images of different angles layer by layer. This process generates more training samples by changing the combinations of different angles of SAR images of the same target [17, 18]. The DCNN network exhibits an improvement in recognition performance under conditions of limited sample size. However, the features extracted from the training data generated by changed combinations contain a significant amount of redundancy, which can lead to model overfitting [19] and a subsequent decline in the algorithm's recognition performance.

Exploring the mapping relationship between target radar cross section (RCS) and SAR image features under multi-view conditions can help to enhance the richness of network feature extraction and improve the network's recognition accuracy of SAR targets. The variation of RCS at different angles of the target is one of the important reasons for the large differences in the SAR images of the same target [20]. This difference in turn leads to the network extracting different features of the same target, making the intra-class distance larger. Changes in target RCS cause changes in the features of the images extracted by the network, and accordingly, the differences in the features of different SAR images can be mapped to some extent to the differences in RCS. Therefore, if the differential features extracted by the network from different angles are fused with the original SAR image features, the new features obtained can map the image features corresponding to the unknown RCS to a certain extent. Therefore, the feature fusion method based on the RCS variation of multi-view target is expected to provide a new research idea for the SAR target recognition method under small sample conditions.

Based on the above analysis, this paper proposes a SAR target recognition method based on a multi-view differential feature fusion recognition network for the problem of SAR target recognition under small sample conditions. The method firstly extracts the original features from different angles of the same target, then extracts the difference between features by feature subtraction, produces richer and finer new features by feature fusion, and finally, achieves accurate recognition of SAR targets under small sample conditions based on the fused features. The main contribution and innovation of this paper is as follows.

(1) A multi-view differential feature fusion recognition network is proposed, which improves SAR target recognition accuracy under small sample conditions by differential features fusion.

(2) Constructing differential features based on target's RCS changes under multiple angles improves the fit of the recognition network to the measured data. The proposed method achieves better recognition results on both Standard Operating Condition (SOC) and Extended Operating Condition (EOC) datasets constructed based on the Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset.

(3) The proposed network has only 5 layers of convolution and does not require preprocessing of SAR images, which reduces the amount of computation and realizes a lightweight network.

The following outlines the structure of this paper. Section 2 provides a detailed description of the network structure of the Multi-view Differential Feature Fusion Network. Section 3 presents and analyzes the experimental results, confirming the superiority of our network. Finally, Section 4 concludes the paper.

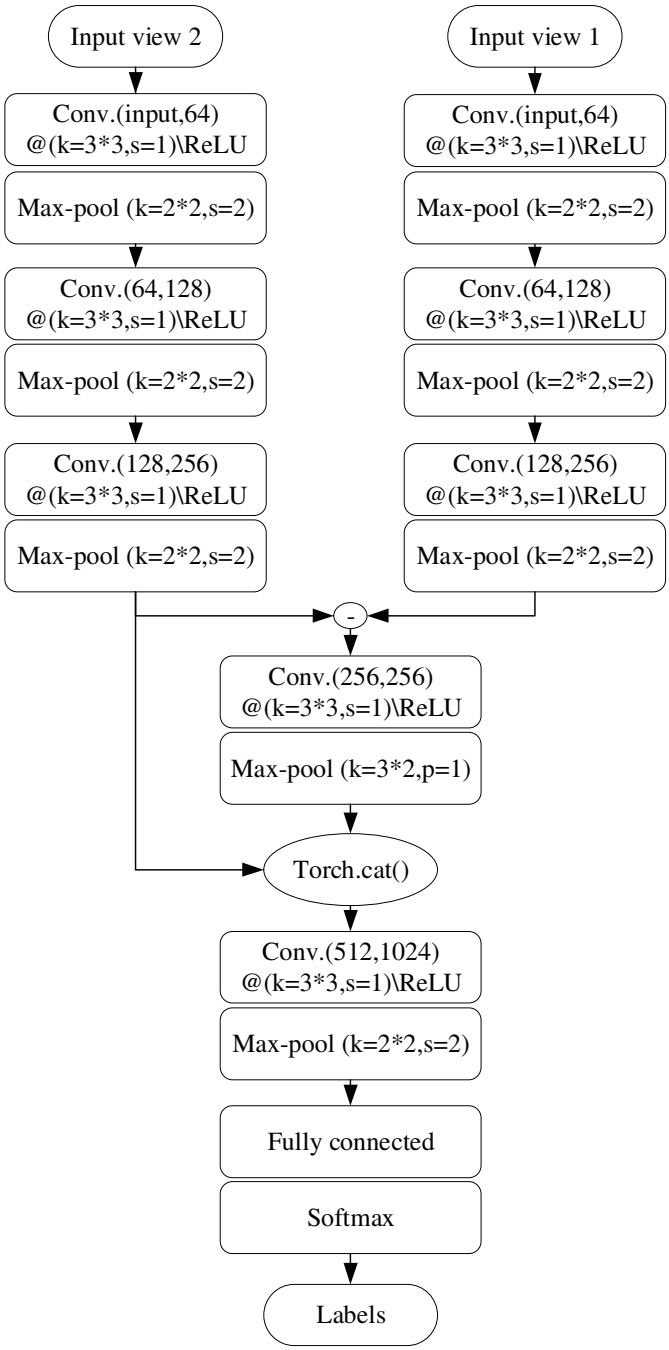

FIGURE 2. The detailed network structure and convolution parameters.

2. PROPOSED MULTI-VIEW DIFFERENTIAL FEATURE FUSION NETWORK

The basic structure of the network is shown in Fig. 1. The proposed method has a dual-input parallel topology, where the inputs are two SAR images of the same target with different angles, Input-view 1 and Input-view 2, respectively. In the feature extraction stage, two input features are extracted simultaneously, namely view 1 feature and view 2 feature, and then the feature differences between the two angles are extracted in depth. Finally, the calculated feature differences and Input-view 2 are fused to obtain a new feature. The new features obtained in this way are able to map the image features corresponding to the unknown RCS to some extent, which in return increases the features learned by the network. Since different combinations of inputs extract different features, repeated combinations of a small amount of data can generate a large amount of usable training data, thus greatly enriching the features learned by the network. At the same time, the parameters of the two input branches of Input-view 1 and Input-view 2 are shared, and the overall network has only five layers of convolution, which greatly reduces the network's complexity and improves the overall training speed.

The detailed structure of the network is shown in Fig. 2. According to the parameters in Fig. 2, it can be seen that in the feature extraction stage, the first three layers of convolution parameters for parallel input are entirely consistent. Therefore, to reduce the complexity of the network, we adopt the strategy of shared convolution. Extract differential features after the third layer convolution and fuse them with Input-view 2.

2.1. Differential Feature Fusion Based on Multi-Angle RCS

From the analyses in Section 2, it can be seen that changes in the RCS of different angles of the same target can cause changes in the target features extracted by the network, and accordingly, the differences in the SAR image features of different angles of the same target can reflect the differences in the target RCS to some extent. The mapping relationship between RCS and image features is shown in Fig. 3 Input-view 1 and Input-view 2 represent the corresponding SAR images when RCS is **a** and **b**, respectively. The new features obtained by extracting the multi-view differential features of the two and fusing them with Input-view 2 can map the SAR image features extracted by the network when RCS equals **c** to a certain extent. Therefore, the permutation of different SAR images can generate a large number of new features for network training and increase the amount of information learned by the network.

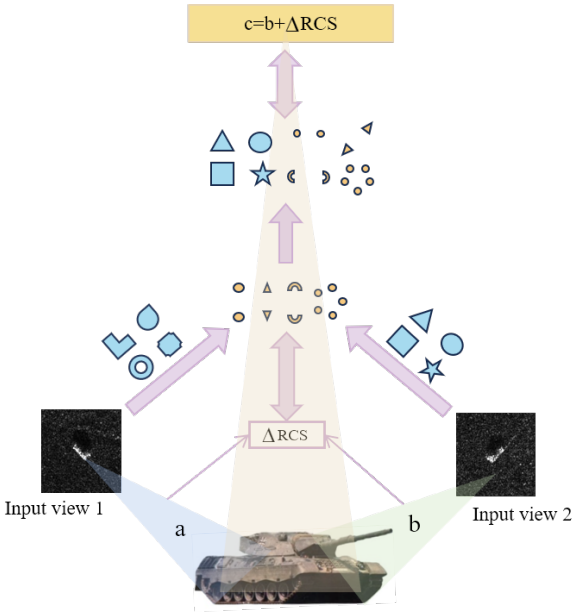

FIGURE 3. Relationship between RCS and features.

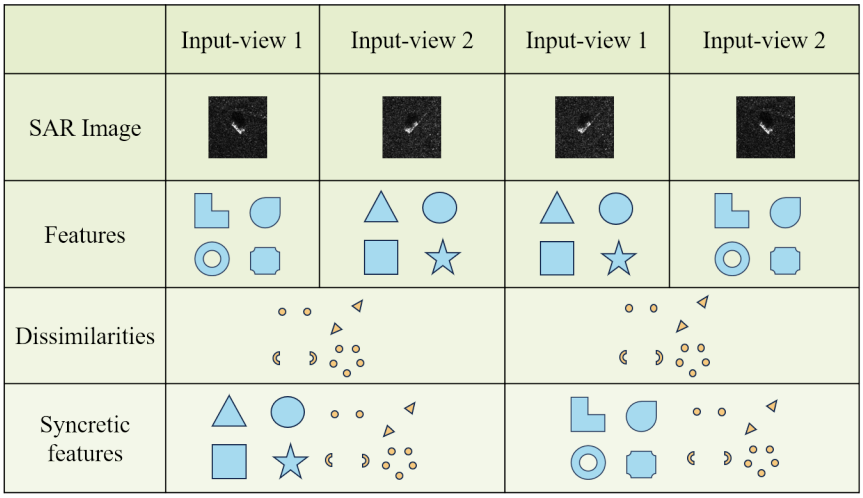

FIGURE 4. Example for data amplification.

Input-view1	Input-view2	label
./MSTAR/train/0/HB03880.jpg	./MSTAR/train/0/HB03955.jpg	0
./MSTAR/train/0/HB03955.jpg	./MSTAR/train/0/HB03880.jpg	0
./MSTAR/train/0/HB03968.jpg	./MSTAR/train/0/HB03889.jpg	0
./MSTAR/train/0/HB03904.jpg	./MSTAR/train/0/HB04002.jpg	0
./MSTAR/train/0/HB04002.jpg	./MSTAR/train/0/HB03901.jpg	0
./MSTAR/train/0/HB03985.jpg	./MSTAR/train/0/HB04027.jpg	0
./MSTAR/train/0/HB04040.jpg	./MSTAR/train/0/HB03904.jpg	0
./MSTAR/train/0/HB04027.jpg	./MSTAR/train/0/HB03960.jpg	0
./MSTAR/train/0/HB03909.jpg	./MSTAR/train/0/HB03796.jpg	0

FIGURE 5. Structure of the label file.

2.2. Data Augmentation Based on Feature Fusion

As shown in Fig. 4, because the network is a two-input parallel structure for Input-view 1 and Input-view 2, respectively, the multi-view differential features extracted from both of them are fused with the original features extracted from Input-view 2 to get the new features, so even if the same two images correspond to different input ports, the final extracted features are not the same. Thus, a large amount of usable training data can be generated by repetitive combination of data. The permutations of the data are made between the same categories, and each image can be combined with all the remaining images of that category.

$$d = mn(n-1) \tag{1}$$

Here d denotes the number of all available training data, m the dataset category, and n the amount of data in each category.

From Equation (1), it can be seen that a large amount of usable training data can be generated by permutations even in the case of very few samples. These permutations can generate a large number of new features for network training, increase the amount of information learned by the network, and improve the model performance.

For the convenience of network training, we achieve the generation of training data by means of the permutations of labels. Our label data is the saving path of each SAR image. With the network training, it inputs the images into the network by accessing the path of each image. By this way, we can easily produce labels, and at the same time generate a large amount of usable training data through the permutations of labels. Fig. 5 shows the structure of the label file, in which each column represents a different input port, and each row represents an input combination. In the red box, the same two images are input into different ports and end up with different features after fusion. Similarly, the blue box is the same SAR image combined with a different SAR image to get a new input pair, and the final extracted features are still different. Thus, by the repetitive combinations of labels, the training data is generated. Theoretically, the number of combinations that can be generated for each category is $n(n-1)$. Considering that too many repetitive combinations may cause feature redundancy, and for the convenience of label production, we extract the save paths of the SAR images for each category by random traversal with a python program and set the traversal parameter to control the number of generated labels, instead of using all the combinations of training data for training the network. Additionally, when the network extracts labels, it uses the combinations of SAR images of non-adjacent angles, which increases the dissimilarities between features and effectively prevents overfit-

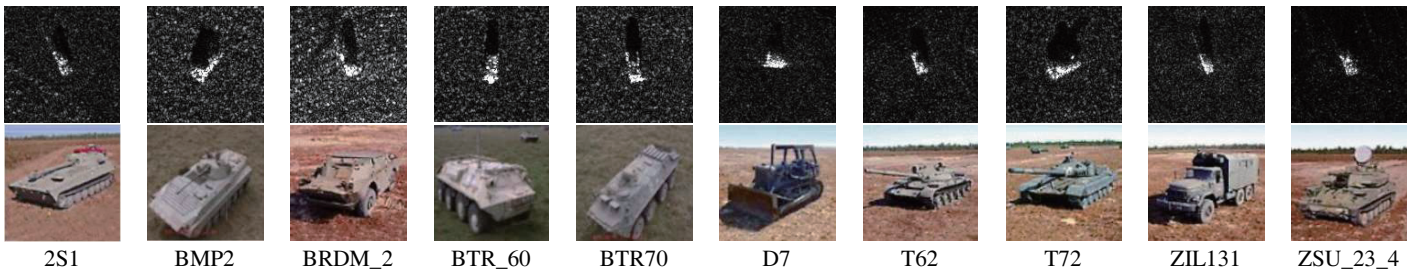

FIGURE 6. Comparison of SAR images with optical images for ten targets in the MSTAR dataset.

ting.

$$P = \alpha nm, \quad \alpha \leq n-1 \tag{2}$$

Here P denotes the number of training data generated; n denotes the number of SAR images per category; and m denotes that there are a total of m categories in the dataset. Since the theoretical amount of data available for each category is $n(n-1)$, the stochastic parameters should satisfy $\alpha \leq n-1$.

2.3. MDFFN for SAR ATR

As mentioned before, we approximate the mapping of SAR image features under unknown RCS by extracting and fusing the different features of different angles, which is implemented as follows.

Considering the mutual mapping relationship between RCS and features, we design a way to extract feature differences by subtracting features. This method obtains the most original feature differences without preprocessing the image, which can maximally avoid the damage by preprocessing operations [21].

$$f_d^l = f_1^l - f_2^l \tag{3}$$

Here f_d^l denotes the multi-view differential features of Input-view 1 and Input-view 2 SAR images after the 3^{rd} layer of convolution. f_1^l and f_2^l denote the original features extracted from Input-view 1 and Input-view 2 SAR images at the 3^{rd} layer of convolution.

Multi-view differential features can represent the feature differences between different angles, and then map the differences between two angles' RCS. In order to obtain the complete features of unknown angles, we need to fuse the multi-view differential features with the extracted original features.

We use feature splicing to fuse the multi-view differential features with the original features of the SAR image. This approach can retain both the original SAR image features and multi-view differential features, and increase the amount of information learned by the network. The generated new features can map the image features under unknown RCS to some extent.

$$f_h = concat(f_2^l, f_d^l) \tag{4}$$

Here, $concat$ represents the feature splicing. f_h represents the spliced features. The new features contain both original features' information and multi-view different features' information. The number of new features generated by combinations of data far exceeds the number of features extracted from the original data. It increases the amount of information learnt by the network.

So far, the proposed multi-view differential feature fusion network has been constructed. Only the SAR images of the same target with different angles are input into the network, and then the differential features are extracted and fused, so that the expansion of the training samples can be realized, and the recognition accuracy of SAR under the condition of small samples can be improved.

3. EXPERIMENTAL RESULTS AND COMPARATIVE ANALYSIS

In this section, network performance is evaluated by constructing a multi-view differential feature fusion network on a measured dataset of SOC and EOC conditions.

3.1. Dataset

In this paper, a Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset is selected for the experiments, which is produced by Sandia National Laboratories, USA, and has been widely used in the field of SAR ground target identification. It consists of a series of $0.3\,\mathrm{m} \times 0.3\,\mathrm{m}$ resolution SAR images acquired by an X-band spotlight SAR sensor. These images contain different types of vehicle targets and clutter. We used ten types of targets in our experiments to evaluate the network performance, including T62 tanks, T72 tanks, 2S1 self-propelled howitzers, ZIL131 cargo trucks, BTR70 armoured transporters, BTR_60 armoured transporters, BRDM_2 armoured reconnaissance vehicles, BMP2 infantry fighting vehicles, ZSU_23_4 self-propelled anti-aircraft artillery pieces and D7 bulldozers. Fig. 6 shows the SAR images as well as the optical images for the ten types of targets included. From the MSTAR dataset, Standard Operating Conditions (SOCs) and Extended Operating Conditions (EOCs) datasets can be constructed. We evaluated the network performance under both conditions. The detailed distribution of data under SOC working conditions in the MSTAR dataset is shown in Table 1.

The proposed method was tested and evaluated on a GPU cluster with an Intel1Xeon1 CPU E5-2698 v4 (2.20 GHz) and one Tesla V100 with a height of 16 GB of memory. The method was implemented using the open-source PyCharm framework, and only one Tesla V100 was used.

TABLE 1. Detailed distribution of MSTAR Dataset at different pitch angles.

target type	BMP2	BRDM_2	BTR_60	BTR70	D7	2S1	T62	T72	ZIL131	ZSU_23_4
Training (17°)	233	298	256	233	299	299	299	232	299	299
Testing (15°)	195	274	195	196	274	274	273	196	277	274

TABLE 2. Results (%) between the proposed algorithm and other current algorithms.

Algorithms	Raw Number						
	20	30	40	60	100	220	all
VGGNet [22]	_	_	_	_	_	_	93.22
MGAN-CNN [23]	85.23	_	90.82	_	_	_	97.81
DCNN [16]	86.08	90.67	_	94.60	98.07	_	99.07
NLCA-Net [24]	_	_	_	_	_	_	99.72
Ours	**94.87**	**96.33**	**97.07**	**98.53**	**99.33**	**99.54**	_

3.2. Experimental Results and Comparative Analysis Under SOC

In this section, we conduct experiments on the recognition performance of ten types of targets in MSTAR data under SOC.

The experimental procedure uses SAR images with a pitch angle of 17° as the training set and SAR images with a pitch angle of 15° as the test set. In the comparative experiment, we randomly selected different numbers (20, 30, 40, 60, 100, 200) of raw samples from each category in the MSTAR dataset for training. To further demonstrate the superiority of the proposed method, we used all data to participate in training in the comparative algorithm. The detailed data volume of all categories in the MSTAR dataset is shown in Table 1. We choose different algorithms for comparative experiments. These methods include: traditional single-input network, 2-view deep convolutional neural network (2V-DCNN), 3-view convolutional neural network (3VDCNN), Multi-grained Attention Network — convolutional neural network (MGAN-CNN) [23], non-local context attention network (NLCA-Net) [24], and other related algorithms.

As shown in Table 2, the experimental results between the proposed algorithm and other current algorithms with different amounts of raw data are demonstrated. "Raw Number" is the number of randomly selected samples from the original MSTAR dataset. "Training Number" is the number of labelled samples which is generated by the algorithm provided by the network that can be used for training by randomly arranging and permuting the original samples. As shown in Fig. 4, the different arrangement orders of SAR images a input to the network lead to different label samples generated. According to the experimental results, it is obvious that the proposed algorithm is effective when the samples are small. Especially when there are 20 original data per category, the recognition rate reaches 94.87%, while the best result of the comparison algorithm is only 86.08%. When the number of samples gradually increases, the recognition effect of the proposed method still has a clear advantage. The recognition rate reaches 99.33% with 100 pieces of raw data per category, which exceeds most of the recognition results obtained by training with all the data. Meanwhile, when the number of samples is sufficient, such as 220 samples in each class, the proposed method still has good recognition effect, and the recognition accuracy reaches 99.54%.

Table 3 shows the results of the comparison experiments between the traditional single-input network and the multi-view differential feature fusion network based on the same number of convolutional layers with different training samples. The proposed network uses only 5 layers of convolution. The network increases the amount of learnt information through new features which are extracted by multi-view differential features' extraction and original features' fusion. Efficient recognition performance is achieved under small sample conditions. The advantages over single input and traditional multi-view fusion recognition algorithms are very obvious in each training condition. At 20 training samples per class, the recognition rate of the proposed method is 94.87%, while the DCNN algorithm can theoretically improve the recognition performance by increasing the number of parallel inputs, but at this time, the recognition performance of the 3V-DCNN is instead smaller than that of the 2V-DCNN. This is because when the amount of data is very small, reusing data to directly carry out feature fusion will produce a large amount of feature redundancy, leading to model overfitting, causing the recognition performance degradation.

On the contrary, the proposed method can avoid feature redundancy, because it extracts feature differences from different angles, and the features extracted from different combinations of data are very different. Moreover, the new features can reflect the image features corresponding to the unknown RCS to a certain extent, which greatly enriches the amount of information learned by the network. At 60 and 100 original samples per type, the recognition rate of the proposed method is 98.53% and 99.33%, which is improved by 3.66% and 4.46%, respectively. The recognition performance improvement is especially obvious compared with the DCNN series algorithms. This is because when there are more original training samples, the proposed method is able to obtain more features by arranging and combining the data, so a small amount of original data can significantly improve the recognition performance of the proposed method. In particular, at 100 original images per type,

TABLE 3. The results of the comparison experiments.

Raw Number	Methods	Recognition rate (%)
20	single input view network	80.08
	2-VDCNN	86.08
	3-VDCNN	84.14
	Ours	**94.87**
60	single input view network	86.60
	2-VDCNN	92.67
	3-VDCNN	94.13
	Ours	**98.53**
100	single input view network	88.91
	2-VDCNN	96.50
	3-VDCNN	97.00
	Ours	**99.33**

TABLE 4. Recognition rate (%) for each target with different samples and different training numbers.

Raw Number	Training number	BMP2	BTR70	T72	2S1	BRDM_2	D7	BTR_60	T62	ZIL131	ZSU_23_4	Average
20	1000	72.82	86.13	67.18	79.59	54.74	77.74	73.99	85.20	99.27	99.64	80.08
	2000	76.00	93.33	87.33	52.00	90.67	94.67	90.67	89.40	89.33	97.33	86.08
	3000	92.00	88.67	69.33	67.33	94.00	89.33	92.00	65.56	88.67	94.67	84.14
	4000	82.67	89.33	100	98	99.33	98	98	87.42	99.33	96.67	94.87
60	8000	78.46	95.62	83.59	92.86	78.83	62.77	81.32	96.94	97.45	99.64	86.60
	16000	97.33	92.00	84.00	83.33	100	98.00	95.33	80.13	96.67	100	92.67
	25000	80.67	88.67	100	94.67	99.33	96.67	98.00	87.42	100	96.00	94.13
	35000	93.33	94	100	99.33	100	99.33	99.33	100	100	100	98.53
100	20000	78.97	87.59	77.95	84.18	81.39	85.77	97.07	95.92	95.26	99.64	88.91
	30000	89.99	94.99	100	94.99	94.99	94.99	100	94.99	100	100	96.50
	50000	90.67	100	98.67	96.67	97.99	99.33	96.67	94.04	97.33	98.67	97.00
	60000	95.33	99.33	100	99.33	100	100	100	99.34	100	100	99.33

the recognition rate is above 99% for all the remaining types, except for BMP2, which has a recognition rate of 95.33%.

The experimental results of the t-distributed stochastic neighbor embedding (t-SNE) [25] visualisation with 20 SAR images for each category of original samples are shown in Fig. 7, where Fig. 7(a), Fig. 7(b), Fig. 7(c) denote the visualisation results of the three methods: single-input, 2V-DCNN and 3V-DCNN, and Fig. 7(d) denotes the visualisation result of the proposed method. According to the clustering results, it is obvious that the proposed method clusters better. Fig. 8 illustrates the confusion matrix of the experimental results of the proposed method for different numbers of original samples.

To show the effectiveness of the proposed method even further, different numbers of training sets as shown in Table 4 are generated by us to train the model. We generated four types of training sets to train the model in different original training samples. It can be found that the recognition performance of the model steadily improves with the increase of the number of training sets. It proves that the new features extracted by the proposed method can indeed represent the image features under unknown RCS to a certain extent, enriching the amount of information learnt by the network and improving the model performance.

3.3. Experimental Results and Comparative Analysis Under EOC

In this section, we will discuss the experimental results of the algorithm in more complex working conditions to verify the robustness of the proposed method. The EOC working conditions acquire more complex datasets, and the difference between the data in the training set and validation set is much more obvious than that of the SOC working conditions, which can well verify the robustness of the algorithm. Specifically, it is divided into three working conditions, EOC-1, EOC-2, and EOC-3. Its data distribution is shown in Table 5.

For this experiment, we randomly selected 20 SAR images from each class of the original data as a training set. A large amount of training data is generated for network training by the proposed method. The experimental results are shown in Table 6. Compared with the experimental results of SOC, the network recognition performance is degraded in EOC. On one

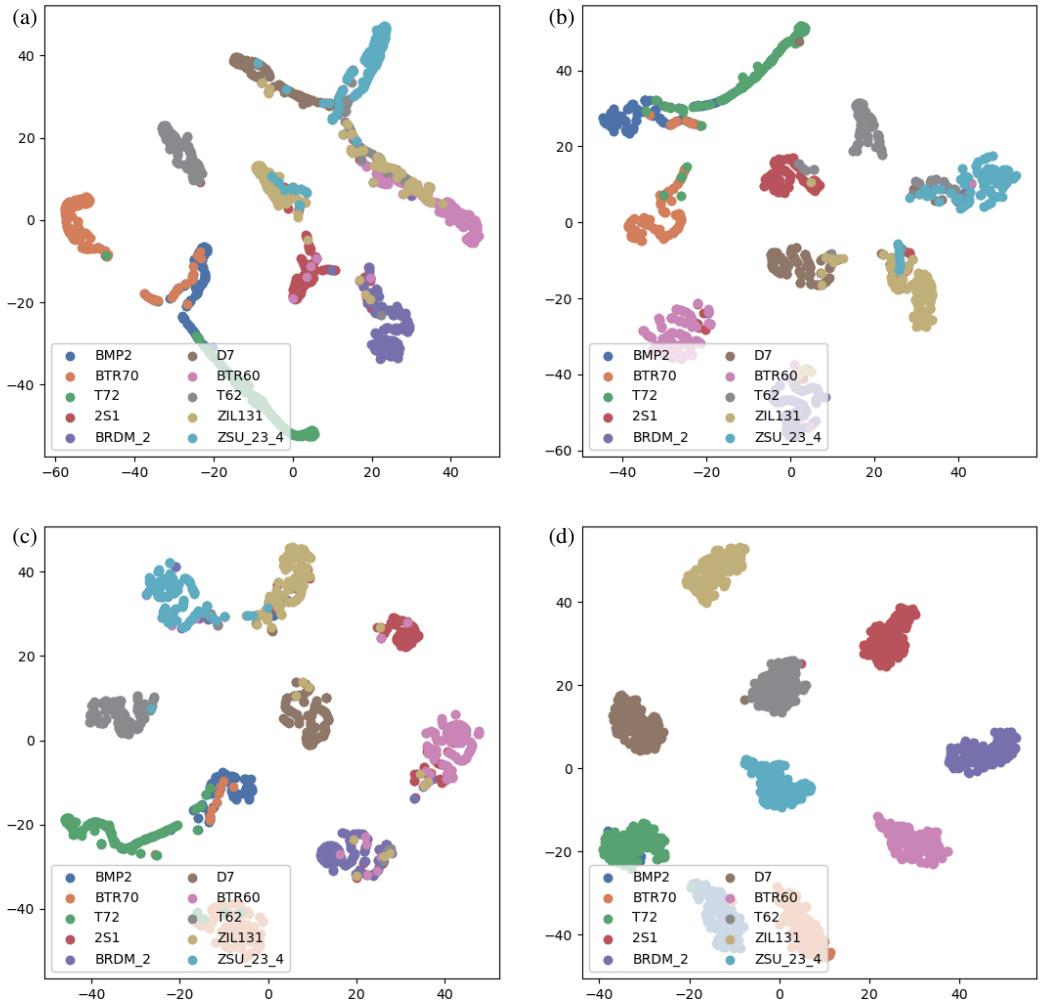

FIGURE 7. Feature visualization of different configurations with 20 each class under MSTAR. (a) Single input view network, (b) 2-VDCNN, (c) 3-VDCNN, (d) Ours.

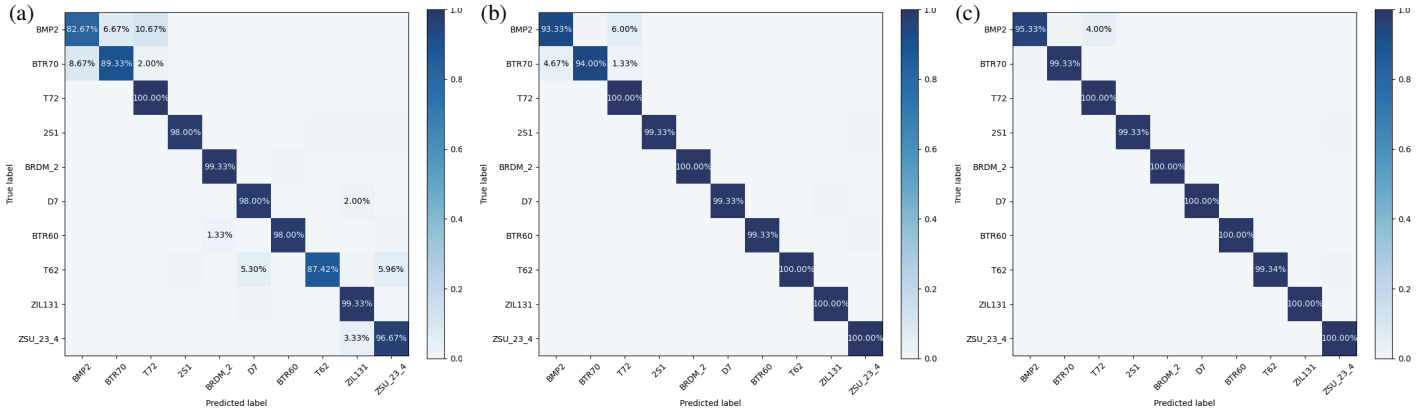

FIGURE 8. Confusion Matrix of network instance under SOC. (a) 20 samples(Recognition rate: 94.87%), (b) 60 samples (Recognition rate: 98.53%), (c) 100 samples(Recognition rate: 99.33%.

hand, it is because the difference between the training set and the validation set is much more obvious in EOC, which brings some difficulties to the network recognition and affects the network recognition performance to some extent. On the other hand, due to the decrease in the number of classifications, the number of training datasets generated by the proposed method

TABLE 5. Training and testing datasets under EOCs.

Conditions	Training			Testing		
EOC-D	Class	Number	Depression Angle	Class	Number	Depression Angle
	2S1	299	17°	2S1	288	30°
	BRDM_2	298		BRDM_2	287	
	T72	232		T72	288	
	ZSU_23_4	299		ZSU_23_4	288	
EOC-V	Class	Number	Depression Angle	Class	Number	Depression Angle
	BMP2	233	17°	T72-SN812	426	17°, 15°
	BRDM2	298		T72-A04	573	
	T72	232		T72-A05	573	
	BTR70	233		T72-A07	573	
				T72-A10	567	
				BMP2-9566	428	
				BMP2-C21	429	
EOC-C	Class	Number	Depression Angle	Class	Number	Depression Angle
	BMP2	233	17°	T72-S7	419	17°, 15°
	BRDM2	298		T72-A32	572	
	BTR70	233		T72-A62	573	
	T72	232		T72-A63	573	
				T72-A64	573	

TABLE 6. Recognition performance (%) under EOCs on MSTAR.

EOC-1				EOC-2				EOC-3			
Class	Raw Number			Class	Raw Number			Class	Raw Number		
	20	60	100		20	60	100		20	60	100
BRDM2-E71	99.33	100	100	BMP2sn-c21	85.33	87.67	96.67	T72sn-s7	97.00	98.67	99.33
2S1-b01	95.00	97.33	99.00	BMP2sn-9566	86.33	89.33	96.00	T72-A32	96.33	98.67	97.33
T72-132	86.67	93.00	96.33	T72sn-812	91.33	98.67	98.67	T72-A62	96.67	97.33	99.33
ZSU234-d08	91.33	96.67	93.33	T72-A04	90.33	96.67	95.33	T72-A63	94.67	96.33	97.67
Average	**93.08**	**96.75**	**97.17**	**Average**	**88.33**	**93.09**	**96.67**	T72-A64	91.67	94.33	94.67
								Average	**95.27**	**97.07**	**97.67**

decreases, leading to a decrease in the features learnt by the model, which further causes a decrease in the recognition performance.

However, from the comparison of the experimental results, the proposed method still performs satisfactorily under EOC. The recognition rate reaches 93.08%, 88.33%, and 95.27% at 20 original images per class. When the amount of original data increases, the recognition performance of the proposed method improves significantly, reaching 97.17%, 96.67%, and 97.67% at 100 images per class. It proves that the proposed method has good robustness and generalization.

4. CONCLUSION

In this paper, a multi-view differential feature fusion network based on convolutional neural network is designed for the problem of SAR image target recognition under small sample conditions. By using the mapping relationship between SAR image features and RCS, multi-view differential feature fusion is realized, which increases a large amount of available training data, avoids complex image preprocessing, reduces computational volume, and overcomes the serious defects of the traditional single-input SAR image recognition network under small sample conditions. The experimental results on the SOC and EOC datasets show that the proposed method has obvious advantages in recognition performance compared with the traditional single-input method and other deep learning network-like algorithms.

ACKNOWLEDGEMENT

The data that support the findings of this study are available from the authors on reasonable request.

REFERENCES

[1] Passah, A., S. N. Sur, A. Abraham, and D. Kandar, "Synthetic Aperture Radar image analysis based on deep learning: A review of a decade of research," *Engineering Applications of Artificial Intelligence*, Vol. 123, 106305, 2023.

[2] Zhao, S., Y. Luo, T. Zhang, W. Guo, and Z. Zhang, "A domain specific knowledge extraction transformer method for multi-source satellite-borne SAR images ship detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 198, 16–29, 2023.

[3] Owirka, G. J., S. M. Verbout, and L. M. Novak, "Template-based SAR ATR performance using different image enhancement techniques," in *Algorithms for Synthetic Aperture Radar Imagery VI*, Vol. 3721, 302–319, 1999.

[4] Ma, C., G. Wen, F. Gao, X. Huang, and X. Yang, "Electromagnetic model based SAR ATR through attributed scatterers," in *Millimetre Wave and Terahertz Sensors and Technology IX*, Vol. 9993, 137–142, 2016.

[5] Sun, Y., Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 43, No. 1, 112–125, 2007.

[6] Zhang, W., Y. Zhu, and Q. Fu, "Deep transfer learning based on generative adversarial networks for SAR target recognition with label limitation," in *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, 1–5, Chongqing, China, Dec. 2019.

[7] Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9, 2015.

[8] Zhang, Y.-P., L. Zhang, L. Kang, H. Wang, Y. Luo, and Q. Zhang, "Space target classification with corrupted HRRP sequences based on temporal–spatial feature aggregation network," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, 1–18, 2023.

[9] Zhou, F., L. Wang, X. Bai, and Y. Hui, "SAR ATR of ground vehicles based on LM-BN-CNN," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 12, 7282–7293, 2018.

[10] El-Darymli, K., E. W. Gill, P. Mcguire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review," *IEEE Access*, Vol. 4, 6014–6058, 2016.

[11] Chen, H., Y. Ji, and D. Evans, "Finding friends and flipping frenemies: Automatic paraphrase dataset augmentation using graph theory," *ArXiv Preprint ArXiv:2011.01856*, 2020.

[12] Zhao, S., Z. Zhang, W. Guo, and Y. Luo, "An automatic ship detection method adapting to different satellites SAR images with feature alignment and compensation loss," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, 1–17, 2022.

[13] Bao, X., Z. Pan, L. Liu, and B. Lei, "SAR image simulation by generative adversarial networks," in *IGARSS 2019 — 2019 IEEE International Geoscience and Remote Sensing Symposium*, 9995–9998, Yokohama, Japan, Jul. 2019.

[14] Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1125–1134, Honolulu, HI, USA, Jul. 2017.

[15] Lu, Q., H. Jiang, G. Li, and W. Ye, "Data augmentation method of SAR image dataset based on wasserstein generative adversarial networks," in *2019 International Conference on Electronic Engineering and Informatics (EEI)*, 488–490, Nanjing, China, Nov. 2019.

[16] Pei, J., Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 4, 2196–2210, 2018.

[17] He, K., 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)*, 770–778, Las Vegas, NV, USA, Jun. 2016.

[18] Chen, S., 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, 4806–4817, 2016.

[19] Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *ArXiv Preprint ArXiv:1312.6229*, 2013.

[20] Charlo, C., S. Méric, F. Sarrazin, E. Richalot, J. Sol, and P. Besnier, "Advanced analysis of radar cross-section measurements in reverberation environment," *Progress In Electromagnetics Research B*, Vol. 104, 51–68, 2024.

[21] Cho, J. H. and C. G. Park, "Multiple feature aggregation using convolutional neural networks for SAR image-based automatic target recognition," *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 12, 1882–1886, 2018.

[22] Simonyan, K., "Very deep convolutional networks for large-scale image recognition," *ArXiv Preprint ArXiv:1409.1556*, 2014.

[23] Fan, F., Y. Feng, and D. Zhao, "Multi-grained attention network for aspect-level sentiment classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3433–3442, Brussels, Belgium, Oct.-Nov. 2018.

[24] Rao, Z., M. He, Y. Dai, Z. Zhu, B. Li, and R. He, "Nlca-net: A non-local context attention network for stereo matching," *APSIPA Transactions on Signal and Information Processing*, Vol. 9, e18, 2020.

[25] Hinton, G. and L. V. D. Maaten, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, Vol. 9, 2579–2605, 2008.