



Communication

FEF-Net: A Deep Learning Approach to Multiview SAR Image Target Recognition

Jifang Pei, Zhiyong Wang, Xueping Sun, Weibo Huo *, Yin Zhang, Yulin Huang, Junjie Wu and Jianyu Yang

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; peijfstudy@126.com or jifangpei@uestc.edu.cn (J.P.); wangzy@std.uestc.edu.cn (Z.W.); 2019210111040@std.uestc.edu.cn (X.S.); yinzhang@uestc.edu.cn (Y.Z.); yulinhuang@uestc.edu.cn (Y.H.); junjie_wu@uestc.edu.cn (J.W.); jyyang@uestc.edu.cn (J.Y.)

* Correspondence: hwbuyi@uestc.edu.cn

Abstract: Synthetic aperture radar (SAR) is an advanced microwave imaging system of great importance. The recognition of real-world targets from SAR images, i.e., automatic target recognition (ATR), is an attractive but challenging issue. The majority of existing SAR ATR methods are designed for single-view SAR images. However, multiview SAR images contain more abundant classification information than single-view SAR images, which benefits automatic target classification and recognition. This paper proposes an end-to-end deep feature extraction and fusion network (FEF-Net) that can effectively exploit recognition information from multiview SAR images and can boost the target recognition performance. The proposed FEF-Net is based on a multiple-input network structure with some distinct and useful learning modules, such as deformable convolution and squeeze-and-excitation (SE). Multiview recognition information can be effectively extracted and fused with these modules. Therefore, excellent multiview SAR target recognition performance can be achieved by the proposed FEF-Net. The superiority of the proposed FEF-Net was validated based on experiments with the moving and stationary target acquisition and recognition (MSTAR) dataset.

Keywords: synthetic aperture radar; multiview; automatic target recognition; deep neural network; feature extraction; feature fusion



Citation: Pei, J.; Wang, Z.; Sun, X.; Huo, W.; Zhang, Y.; Huang, Y.; Wu, J.; Yang, J. FEF-Net: A Deep Learning Approach to Multiview SAR Image Target Recognition. *Remote Sens.* **2021**, *13*, 3493. <https://doi.org/10.3390/rs13173493>

Academic Editors: Addabbo Pia, Luca Pallotta and Christos Ilioudis

Received: 21 July 2021

Accepted: 31 August 2021

Published: 2 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Synthetic aperture radar (SAR) [1–3] is an important modern microwave sensor system, with powerful capabilities, including high-resolution imaging, day-and-night use, and all-weather operation. Those qualities make it superior to other sensors, such as infrared and optical sensors, for some applications. With advances in SAR signal processing and imaging performance, people have been paying more attention to classifying or recognizing targets of interest from SAR images. Therefore, automatic target classification or recognition (ATR) has become an attractive but challenging problem in SAR research and application areas [4–8].

Generally, an SAR ATR system discovers regions of interest containing potential targets from the SAR image [9–11] and efficiently assigns those targets reliable and intelligent category labels [12]. Over the years, researchers focused on this field have proposed many novel SAR ATR approaches [7,8]. Many SAR ATR methods or algorithms have also been employed in the past few decades, such as support vector machine (SVM) [13], conditional Gaussian model (CGM) [14], adaptive boosting (AdaBoost) [15], sparse representation [16], and iterative graph thickening (IGT) [17].

The SAR ATR methods mentioned above generally perform well in applications. Nevertheless, many of these methods must often extract handcrafted features from SAR targets, so sophisticated algorithms for ATR must be predesigned. With the rise in machine learning theory in recent years, ATR methods based on deep learning have quickly advanced [18,19]. They can spontaneously learn hierarchical features from the input data and can achieve

remarkable performance in complex ATR tasks. Many novel works using deep neural networks have proven to be powerful tools for SAR ATR [20–23].

Most SAR ATR methods focus on single-input SAR images. However, modern SAR sensors can obtain high-resolution target images from different views in practice. Researchers have indicated that SAR ATR benefits from multiview measurements, since multiview SAR images contain more abundant classification information than single-view images [24]. Thus, studies of multiview SAR ATR methods have started in recent years, achieving good recognition results [25–28].

Although multiview SAR images have more classification information and show great potential for ATR, two important problems should be solved for ATR performance improvements. SAR target images are sensitive to their imaging views, and the same target often has geometric variations in multiview SAR images, such as orientation and shape variations. Hence, the first challenge is effectively extracting the inherent classification features from each view of the SAR image while accommodating their geometric variations. To further exploit the multiview classification information, effective fusion means should be employed to integrate extracted features from multiple views. Therefore, a valid multiview SAR ATR approach should be able to extract the inherent classification features from each view and to fuse these features effectively. Meanwhile, we hope that the processes of feature extraction and fusion can be carried out spontaneously and without much manual intervention. Hence, an end-to-end deep learning network with feature extraction and fusion modules is a perfect choice. It should extract and fuse useful features of the multiview SAR images through network construction and sample training, and thereby achieve superior ATR results.

This paper proposes an end-to-end deep feature extraction and fusion network (FEF-Net) to address these two problems and to improve the multiview SAR ATR performance. Its network architecture is based on a type of multiple input topology for multiview SAR ATR. Some specific modules, such as deformable convolution and squeeze-and-excitation (SE), are embedded in this network. The deformable convolutional layer can extract inherent classification information and can accommodate the geometric variations of SAR target well, whereas the SE module fuses the features from the multiview SAR images together. Thus, these two problems in multiview SAR ATR, classification feature extraction and fusion of input multiview SAR images, can be effectively resolved with FEF-Net. Therefore, the proposed network can exploit classification information from multiview SAR images and can achieve satisfactory ATR performance.

The main contributions of FEF-Net compared with existing SAR ATR methods are the following: (1) We designed a new deep neural network based on a multiple-input topological structure that can significantly improve SAR ATR performance; (2) the helpful classification features of the input multiview SAR images can be extracted and fused thoroughly through the implantation of distinct network modules; and (3) the proposed FEF-Net achieves excellent recognition performance compared with the available SAR ATR methods.

The remainder of this paper is organized as follows: Multiview SAR ATR is modeled and formulated in Section 2. Section 2 also details the proposed FEF-Net for multiview SAR ATR. The ATR performance of FEF-Net is evaluated in Section 3, and Section 4 provides the conclusion.

2. Materials and Methods

2.1. Problem Formulation

Figure 1 shows the multiview SAR ATR geometric model of a ground target. In a practical multiview SAR ATR pattern, the SAR system receives its returns and obtains multiview images of the ground target from different aspects and depressions. For simplicity, the depression is set as a constant here. When the view interval θ and view number k are provided, SAR can collect the ground target images in multiview imaging mode. Using

these multiview SAR images, more classification information can be obtained than from the single-view pattern.

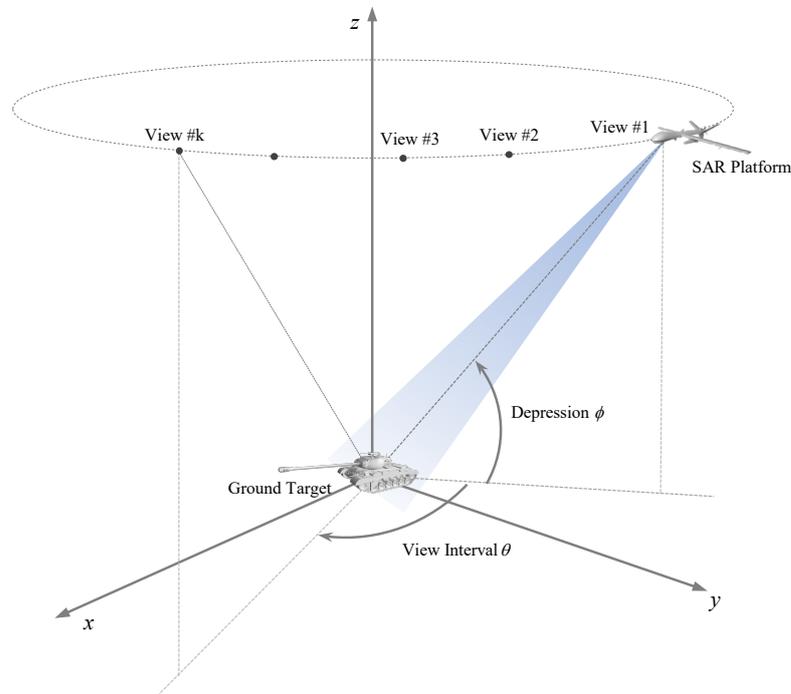


Figure 1. Multiview SAR ATR geometric model of a ground target.

Hence, the multiview SAR target recognition problem requires a valid multiple input classifier to determine the most probable classified label for the interested target, which can be formulated as follows:

$$f : f(\mathbf{x}_{\#1}, \mathbf{x}_{\#2}, \dots, \mathbf{x}_{\#k}) \rightarrow y_i \in \mathcal{Y} \quad (1)$$

In Equation (1), f is the classifier with multiview SAR images as the input, y_i is the assigned target label, \mathcal{Y} is the class label set, and $\mathbf{x}_{\#k}$ is the k th view of the SAR image target with its aspect angle $\varphi(\mathbf{x}_{\#k})$, which satisfies the following conditions.

$$\begin{cases} \varphi(\mathbf{x}_{\#1}) < \varphi(\mathbf{x}_{\#2}) < \dots < \varphi(\mathbf{x}_{\#k}) \\ |\varphi(\mathbf{x}_{\#1}) - \varphi(\mathbf{x}_{\#k})| \leq \theta \end{cases} \quad (2)$$

or

$$\begin{cases} \varphi(\mathbf{x}_{\#1}) > \varphi(\mathbf{x}_{\#2}) > \dots > \varphi(\mathbf{x}_{\#k}) \\ |\varphi(\mathbf{x}_{\#1}) - \varphi(\mathbf{x}_{\#k})| \leq \theta \end{cases} \quad (3)$$

FEF-Net was designed to solve the ATR problem with multiview SAR images based on this formulated model. The FEF-Net method is explained in the following subsection.

2.2. Proposed Method

This subsection proposes FEF-Net to solve these two difficulties and to improve the performance of multiview SAR ATR. The architecture of the FEF-Net for multiview SAR ATR is provided, along with details on specific modules of the network.

2.2.1. Network Framework

The basic framework of the FEF-Net instance with three inputs is shown in Figure 2. This deep neural network is based on a multiple-input topological structure. The inputs are merged into a certain layer, extracting and fusing the classification information from the multiview SAR images.

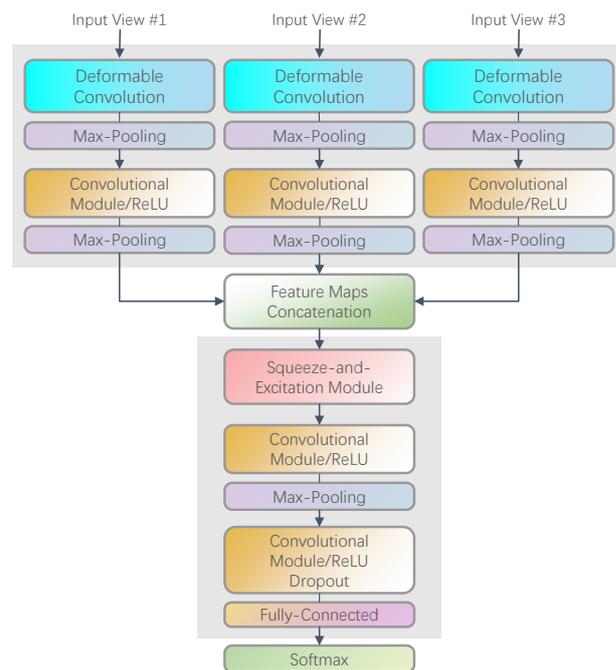


Figure 2. The basic architecture of FEF-Net for multiview SAR ATR.

As mentioned in Section 1, the key point in multiview SAR ATR is an effective extraction and fusion classification feature. Hence, the proposed FEF-Net begins with a deformable convolutional layer in each branch to extract the inherent classification feature from each view and to accommodate the geometric variations in the SAR target. Alternate pooling and convolutional layers are present within each branch to further extract the features from each view and to reduce the feature dimensions. After the feature extraction from each view, the three branch feature maps are concatenated. These merged feature maps are linked to an SE network module to further recalibrate the feature responses and to fuse the concatenated features of the multiview SAR images together. Finally, the FEF-Net instance ends with a fully connected layer, and the softmax classifier performs the recognition decision.

From the basic architecture of the instance, we can see that the proposed FEF-Net can effectively extract and fuse the classification information from the input multiview SAR images, which benefits multiview SAR ATR. Specific modules in the proposed network are provided in the following discussion.

2.2.2. Deformable Convolution

The convolution operation is inspired by the process of the biological neuron in the visual cortex [29]. Supposing that the grid Ω represents the receptive field size and dilation, the convolution operation can be written as follows:

$$\mathbf{z}(\mathbf{p}) = \sum_{\mathbf{p}_n \in \Omega} \mathbf{w}(\mathbf{p}_n) \cdot \mathbf{a}(\mathbf{p} + \mathbf{p}_n) \quad (4)$$

where $\mathbf{z}(\mathbf{p})$ represents the intensity of each location \mathbf{p} on the output feature map \mathbf{z} , \mathbf{w} denotes the convolution kernel, \mathbf{p}_n enumerates the locations in Ω , and \mathbf{a} is the input feature map. An activation function, such as the rectified linear unit (ReLU), follows the convolution to enhance the nonlinear representation of the network.

The diagram for deformable convolution is shown in Figure 3a, which can be formulated as follows:

$$\mathbf{z}(\mathbf{p}) = \sum_{\mathbf{p}_n \in \Omega} \mathbf{w}(\mathbf{p}_n) \cdot \mathbf{a}(\mathbf{p} + \mathbf{p}_n + \Delta\mathbf{p}_n) \quad (5)$$

where $\Delta \mathbf{p}_n$ represents the additional offsets learned using a convolutional operation over the same input feature. Thus, the sampling of the deformable convolution is from the irregular and offset locations $\mathbf{p}_n + \Delta \mathbf{p}_n$.

Since the offset $\Delta \mathbf{p}_n$ is typically fractional, the deformable convolution should be performed based on the interpolation:

$$\mathbf{a}(\mathbf{p}') = \sum_{\mathbf{q}} B(\mathbf{q}, \mathbf{p}') \cdot \mathbf{a}(\mathbf{q}) \quad (6)$$

where \mathbf{q} enumerates all integral spatial locations in \mathbf{a} , $\mathbf{p}' = \mathbf{p} + \mathbf{p}_n + \Delta \mathbf{p}_n$, and $B(\cdot, \cdot)$ denotes the interpolation kernel.

By augmenting the spatial sampling locations with additional offsets, the deformable convolution can enhance the modeling of targets' geometric variations and can effectively extract the inherent features of the SAR images.

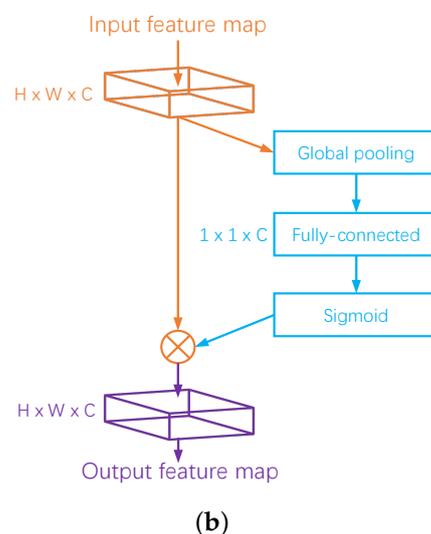
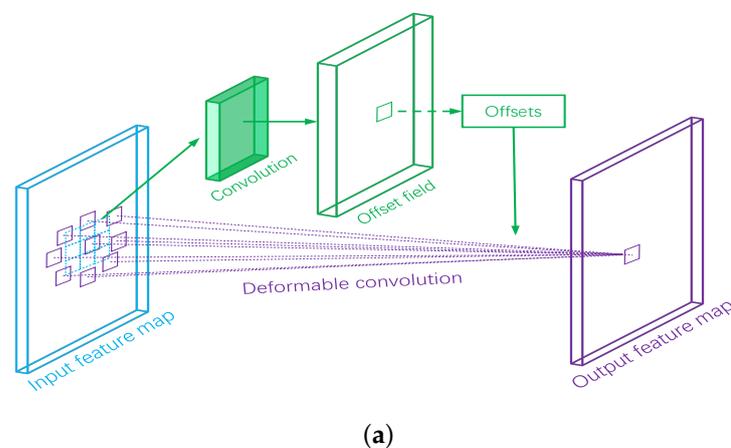


Figure 3. A basic architectural diagram of deformable convolution and SE. (a) Deformable convolution. (b) SE.

2.2.3. SE Module

The classification features extracted from multiple views are different. Thus, we employ an SE module [30] to adaptively recalibrate and fuse those concatenated feature responses from the multiview SAR images. Figure 3b shows a basic block diagram of the SE module. Let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_C]$ denote the input feature maps of the SE module and

$\mathbf{a}_l \in R^{H \times W}$, $l = 1, 2, \dots, C$. The SE module squeezes the global spatial information of the input into a channel descriptor $\mathbf{d} \in R^C$ with a global average pooling operation:

$$d_l = \frac{1}{H \times W} \sum_i \sum_j \mathbf{a}_l(i, j) \quad (7)$$

A fully connected layer is employed to exploit the aggregated information in the squeeze step and adaptively learn the recalibration, formulated as follows:

$$\mathbf{s} = \sigma(\mathbf{W}\mathbf{d} + \mathbf{b}) \quad (8)$$

where $\mathbf{s} = [s_1, s_2, \dots, s_C]$. \mathbf{W} and \mathbf{b} are the weight matrix and bias of the fully connected layer, which are the trainable parameters and can be computed by the network training method [31]. The sigmoid activation is denoted by $\sigma(\cdot)$.

The final fused feature $\mathbf{A}' = [\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_C]$ of SE is obtained by recalibrating the input feature maps \mathbf{A} as follows:

$$\mathbf{a}'_l = F(\mathbf{a}_l, s_l) = s_l \cdot \mathbf{a}_l \quad (9)$$

Through dynamic recalibration and fusion of the features from different views, the SE module can effectively help improve the feature discriminability and ATR performance of FEF-Net.

2.2.4. Other Modules

Other helpful modules or operations, such as pooling, dropout, and softmax, are also necessary for FEF-Net. As an important module in FEF-Net, the pooling layer can extract the prominent features from the input feature map while reducing its dimensions. Here, we use a max-pooling operation in the proposed neural network.

Dropout is an operation widely used to reduce the overfitting of the neural network. It enhances the robustness of the network's learning ability with random active neuron combinations. Dropout is used after the last convolutional layer in FEF-Net to increase the generalization.

After all of the features of the multiview SAR images are extracted and fused, the feature maps are transformed and connected to a fully connected layer. Finally, the softmax classifier [32] provides the accurately classified attributes of the target, as follows:

$$p(y_i | \mathbf{z}^{(L)}) = \frac{\exp(z_i^{(L)})}{\sum_{j=1}^K \exp(z_j^{(L)})} \quad (10)$$

where $\mathbf{z}^{(L)}$ is the input feature vector to the softmax classifier, and K denotes the class number.

2.2.5. Loss Function and Network Training

The loss function used in FEF-Net is the cross-entropy loss [33]:

$$L = - \sum_{i=1}^K y_i \log p(y_i | \mathbf{z}^{(L)}) \quad (11)$$

The training process of FEF-Net is similar to that of a standard SAR ATR neural network, although it has a complex network structure. The back propagation algorithm can be used to calculate the gradients and update the network parameters to effectively train the network.

3. Results and Discussion

3.1. Dataset

We selected raw SAR images from the moving and stationary target acquisition and recognition (MSTAR) dataset [34] to assess the proposed FEF-Net ATR performance for our experiments. The MSTAR program collected a significant quantity of SAR images to evaluate the performances of advanced SAR ATR methods. The MSTAR dataset includes a large number of $0.3 \text{ m} \times 0.3 \text{ m}$ resolution SAR images processed with an X-band spotlight SAR sensor. Ten classes of targets were used in this experiment. The optical images and the corresponding SAR images of these targets are shown in Figure 4.

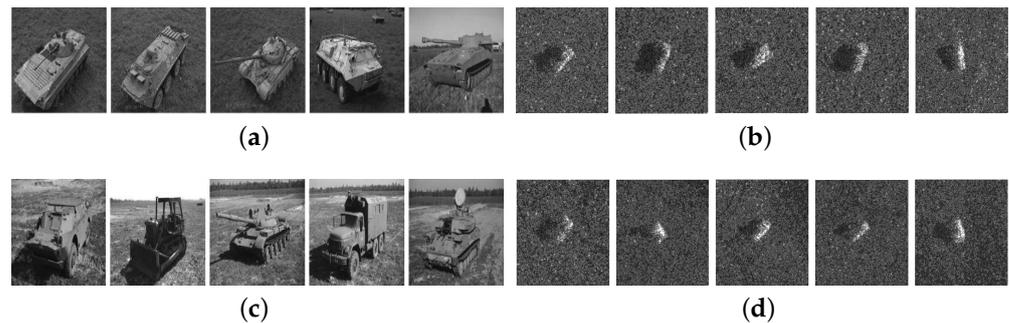


Figure 4. Optical and SAR images of targets in MSTAR dataset. Optical and SAR images for targets of (a,b) BMP2, BTR70, T72, BTR60, and 2S1; (c,d) BRDM2, D7, T62, ZIL131, and ZSU23/4.

Only part of the raw SAR images were selected from these ten class targets with a depression of 17° to generate multiview SAR image samples for training the network in this experiment. The azimuth angles of the selected images for each class all ranged $0^\circ - 360^\circ$. All raw images in the dataset with a depression of 15° were used to generate testing multiview SAR images. The usage of raw SAR images is listed in Table 1. Additionally, the gray enhancement method with a power function [35] was employed to enhance the scattering information of the SAR target images.

Table 1. Raw SAR image selection from the MSTAR dataset.

Training		Testing	
Target Types	Raw Images	Target Types	Raw Images
BMP2sn-9563	78	BMP2sn-9563	195
BTR70	78	BTR70	196
T72sn-132	78	T72sn-132	196
BTR60	86	BTR60	195
2S1	100	2S1	274
BRDM2	100	BRDM2	274
D7	100	D7	274
T62	100	T62	273
ZIL131	100	ZIL131	274
ZSU23/4	100	ZSU23/4	274

The view interval θ was set as 45° . The data augmentation method [27] generated many multiview training samples from the selected raw SAR images. There were 48,764 multiview SAR image samples with a depression of 17° for deep network training. The samples were randomly selected from the generated multiview SAR images with a depression of 15° for testing.

3.2. Network Configuration

The input SAR image size for the network instance was 80×80 , and the probability of dropout was set to 0.5 during the training phase. Table 2 lists the hyper-parameters of

the FEF-Net instance in our experiment, which were determined by statistical validation and trials.

Table 2. FEF-Net instance configurations. Convolutional layers are represented as Conv. and their hyper-parameters denote as (number of feature maps)@(kernel size in convolution). “ $L_a\text{-}\#b$ ” represents the b th branch of a th layer. “WS,” “SS,” and “NN” denote window size, stride size, and number of neurons in network, respectively.

Layer	$L1_{\#1,\#2,\#3}$	$L2_{\#1,\#2,\#3}$	$L3_{\#1,\#2,\#3}$	$L4_{\#1,\#2,\#3}$	L5	L6	L7	L8	L9	L10
Module	Deformable Conv.	Max-Pool	Conv.	Max-Pool	SE	Conv.	Max-Pool	Conv.	Fully-Connected	Softmax
Configuration	8@5 × 5 SS: 1 × 1	WS: 2 × 2 SS: 2 × 2	16@6 × 6 SS: 1 × 1	WS: 2 × 2 SS: 2 × 2	NN: 48	96@5 × 5 SS: 1 × 1	WS: 2 × 2 SS: 2 × 2	128@4 × 4 SS: 1 × 1	NN: 64	NN: 10

3.3. Performance Analysis

Table 3 shows the proposed FEF-Net recognition result with a confusion matrix. The rows of the matrix are ground truths, and the columns are the predicted class labels. Each element in the confusion matrix denotes the recognition rate of FEF-Net for a specific target class.

Table 3 shows that the recognition rate of the proposed FEF-Net was higher than 99.00% in ten classes of the ATR problem. We can infer from this experimental results that the multiview SAR images of the same target contained large amounts of classification information. The proposed FEF-Net can effectively extract and fuse the classification features of the input multiview SAR images using only a few raw data for generating training samples. Therefore, it can exploit recognition information well from multiview SAR images and can achieve an excellent ATR performance.

Table 3. The confusion matrix of the proposed network instance (recognition rate: 99.31%).

Class	BMP2sn-9563	BTR70	T72sn-132	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU23/4
BMP2sn-9563	99.10	0.00	0.10	0.00	0.65	0.00	0.00	0.15	0.00	0.00
BTR70	0.00	99.95	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
T72sn-132	0.75	0.00	99.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BTR60	0.25	0.65	0.00	98.10	0.10	0.65	0.00	0.25	0.00	0.00
2S1	0.00	0.05	0.00	0.00	99.85	0.10	0.00	0.00	0.00	0.00
BRDM2	0.00	0.00	0.00	0.00	0.25	98.85	0.00	0.00	0.90	0.00
D7	0.00	0.00	0.00	0.00	0.00	0.00	98.90	0.00	1.05	0.05
T62	0.00	0.00	0.00	0.10	0.05	0.00	0.00	99.85	0.00	0.00
ZIL131	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.00	99.40	0.00
ZSU23/4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	99.90

To visually show the classification capabilities of FEF-Net, some of the input multiview SAR samples and their output vectors in the fully connected layer were mapped into two-dimensional Euclidean space by the t-distributed stochastic neighbor embedding (t-SNE) algorithm [36], as shown in Figure 5.

Figure 5a shows the input multiview SAR samples, and Figure 5b illustrates the corresponding outputs. We can observe that the visualization results of the original samples are mixed in Figure 5a and were difficult to classify. After being processed by FEF-Net, the samples with the same class label became closer, whereas the different class samples tended to end up far away from each other, as shown in Figure 5b. That allowed for easier classification and led to effective recognition results.

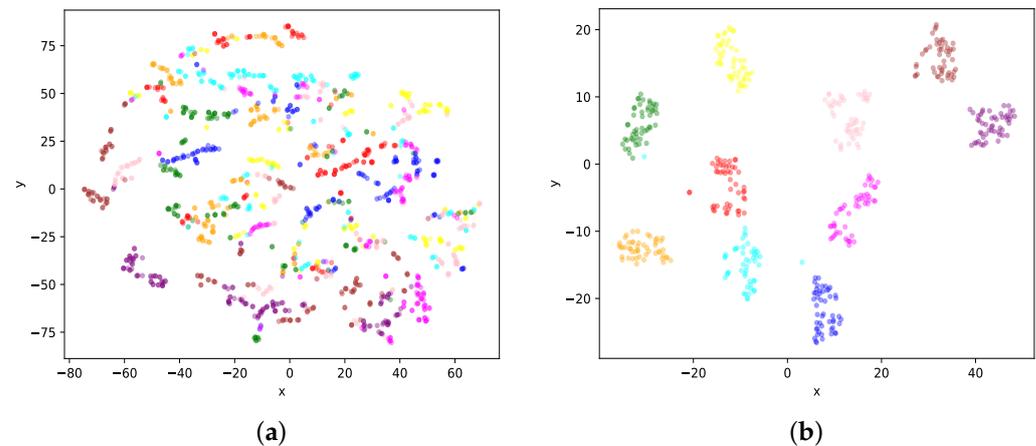


Figure 5. Visualization of classification results of FEF-Net. Points with the same color belong to the same target class. (a) Input multiview samples and (b) corresponding output.

Although the proposed network instance has only three branches in Figure 2, the architecture of FEF-Net is flexible and can have a different number of input branches. We conducted a group of experiments to test the recognition performances of FEF-Net instances with different views. Similarly to the previous experiment, some of the raw SAR images were selected to generate multiview samples for training the networks, and the testing samples were randomly selected for recognition result evaluation. Table 4 shows the raw SAR images, generated training samples, and recognition results of FEF-Net instances with two, three, and four input views.

Table 4. Recognition performances of network instances from different views.

Network Instances	Raw SAR Images	Generated Training Samples	Recognition Rates
2-Views	1377	21,834	98.42%
3-Views	920	48,764	99.31%
4-Views	690	43,533	99.34%

We can see from the experimental results that the recognition rates of FEF-Nets with two, three, and four views were all higher than 98.00%, but only using a few raw data for training sample generation. As the number of input views increased, the recognition rate rose as well, and could reach more than 99.30%. These experimental results indicate the flexibility and potential applications of FEF-Net for SAR ATR tasks with different input views.

We compared the target recognition performance of FEF-Net with the performances of five other methods, including adaptive boosting (AdaBoost) [15], iterative graph thickening (IGT) [17], conditional Gaussian model (CGM) [14], multiview deep convolutional neural network (MDCNN) [27], and sparse representation-based classification (SRC) [25], which are all typical or recently published methods for SAR ATR.

The recognition rate for each ATR method is shown in Figure 6. Although the recognition rates of all the methods were more than 92.00%, their performances were different. The comparisons indicate that FEF-Net had superior recognition performance compared with the other five SAR ATR methods, demonstrating the reasonability and validity of FEF-Net.

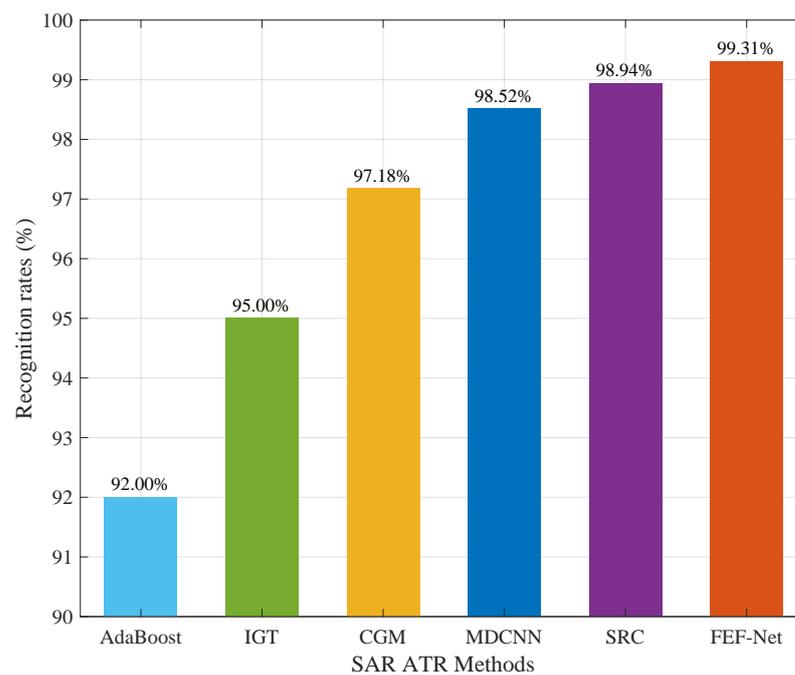


Figure 6. Recognition performances of various methods.

4. Conclusions

Inherent classification feature extraction from each view and multiview feature fusion are two important issues for improving the performance of multiview SAR ATR. We presented a novel ATR approach based on FEF-Net with multiview SAR images. FEF-Net was designed with a multiple-input topological structure, including specific modules, such as deformable convolution and SE, and it has the capability of learning useful classification information for multiview SAR images. Thus, the two key problems, classification feature extraction and fusion, are solved with the proposed FEF-Net. Extensive experiments on the MSTAR dataset were conducted, which showed that the proposed multiview FEF-Net can achieve excellent recognition performance. Its top recognition rate was more than 99.00% in a ten class problem. Additionally, it achieved superior recognition performance compared with the existing SAR ATR methods.

The proposed method attained satisfactory recognition results in SAR ATR because of its effective extraction and fusion of classification features. Although we used helpful learning modules in our network for this study, other promising feature extraction and fusion methods may also work, such as the spatial transformer technique and self-attention mechanism. Hence, these alternative methods are important issues worth studying in subsequent research.

Author Contributions: Conceptualization, J.P.; methodology, J.P. and Z.W.; software, J.P., X.S. and Z.W.; validation, X.S., W.H., Y.H., Y.Z., and J.W.; formal analysis, W.H. and J.Y.; investigation, W.H., Y.H., Y.Z., and J.W.; resources, X.S., J.W., and J.Y.; writing—original draft preparation, J.P.; visualization, Y.Z., W.H., and Z.W.; supervision, Y.H. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant numbers 61901091 and 61901090.

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

References

1. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [\[CrossRef\]](#)
2. Brown, W.M. Synthetic Aperture Radar. *IEEE Trans. Aerosp. Electron. Syst.* **1967**, *AES-3*, 217–229. [\[CrossRef\]](#)
3. Doerry, A.W.; Dickey, F.M. Synthetic aperture radar. *Opt. Photonics News* **2004**, *15*, 28–33. [\[CrossRef\]](#)
4. Blacknell, D.; Griffiths, H. *Radar Automatic Target Recognition (ATR) and Non-Cooperative Target Recognition (NCTR)*; The Institution of Engineering and Technology: Stevenage, UK, 2013.
5. Bhanu, B. Automatic Target Recognition: State of the Art Survey. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *AES-22*, 364–379. [\[CrossRef\]](#)
6. Mishra, A.K.; Mulgrew, B. Automatic target recognition. *Encycl. Aerosp. Eng.* **2010**, 1–8. [\[CrossRef\]](#)
7. El-Darymli, K.; Gill, E.W.; McGuire, P.; Power, D.; Moloney, C. Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review. *IEEE Access* **2016**, *4*, 6014–6058. [\[CrossRef\]](#)
8. Kechagias-Stamatis, O.; Aouf, N. Automatic Target Recognition on Synthetic Aperture Radar Imagery: A Survey. *IEEE Aerosp. Electron. Syst. Mag.* **2021**, *36*, 56–81. [\[CrossRef\]](#)
9. El-Darymli, K.; McGuire, P.; Power, D.; Moloney, C. Target detection in synthetic aperture radar imagery: A state-of-the-art survey. *J. Appl. Remote Sens.* **2013**, *7*, 071598. [\[CrossRef\]](#)
10. Kreithen, D.E.; Halversen, S.D.; Owirka, G.J. Discriminating targets from clutter. *Lincoln Lab. J.* **1993**, *6*, 25–52.
11. Gao, G. An Improved Scheme for Target Discrimination in High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 277–294. [\[CrossRef\]](#)
12. Novak, L.M.; Owirka, G.J.; Weaver, A.L. Automatic target recognition using enhanced resolution SAR data. *IEEE Trans. Aerosp. Electron. Syst.* **1999**, *35*, 157–175. [\[CrossRef\]](#)
13. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 643–654. [\[CrossRef\]](#)
14. O’Sullivan, J.A.; DeVore, M.D.; Kedia, V.; Miller, M.I. SAR ATR performance using a conditionally Gaussian model. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 91–108. [\[CrossRef\]](#)
15. Sun, Y.; Liu, Z.; Todorovic, S.; Li, J. Adaptive boosting for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 112–125. [\[CrossRef\]](#)
16. Zhang, H.; Nasrabadi, N.M.; Huang, T.S.; Zhang, Y. Joint sparse representation based automatic target recognition in SAR images. *SPIE Defense Secur. Sens. Int. Soc. Opt. Photonics* **2011**, *8051*, 805112.
17. Srinivas, U.; Monga, V.; Raj, R.G. SAR Automatic Target Recognition Using Discriminative Graphical Models. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 591–606. [\[CrossRef\]](#)
18. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
19. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [\[CrossRef\]](#)
20. Chen, S.; Wang, H.; Xu, F.; Jin, Y.Q. Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4806–4817. [\[CrossRef\]](#)
21. Wagner, S.A. SAR ATR by a combination of convolutional neural network and support vector machines. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 2861–2872. [\[CrossRef\]](#)
22. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional Neural Network With Data Augmentation for SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368.
23. Zhou, F.; Wang, L.; Bai, X.; Hui, Y. SAR ATR of Ground Vehicles Based on LM-BN-CNN. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7282–7293.
24. Brendel, G.F.; Horowitz, L.L. Benefits of aspect diversity for SAR ATR: Fundamental and experimental results. In *AeroSense 2000*; International Society for Optics and Photonics: Orlando, FL, USA, 2000; pp. 567–578.
25. Ding, B.; Wen, G. Exploiting multi-view SAR images for robust target recognition. *Remote Sens.* **2017**, *9*, 1150.
26. Zhang, F.; Hu, C.; Yin, Q.; Li, W.; Li, H.; Hong, W. Multi-Aspect-Aware Bidirectional LSTM Networks for Synthetic Aperture Radar Target Recognition. *IEEE Access* **2017**, *5*, 26880–26891. [\[CrossRef\]](#)
27. Pei, J.; Huang, Y.; Huo, W.; Zhang, Y.; Yang, J.; Yeo, T. SAR Automatic Target Recognition Based on Multiview Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2196–2210. [\[CrossRef\]](#)
28. Clemente, C.; Pallotta, L.; Proudler, I.; De Maio, A.; Soraghan, J.J.; Farina, A. Pseudo-Zernike-based multi-pass automatic target recognition from multi-channel synthetic aperture radar. *IET Radar Sonar Navigat.* **2015**, *9*, 457–466.
29. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference On Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
31. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
32. Trevor Hastie, R.T.; Friedman, J. *The Elements of Statistical Learning; Data Mining, Inference and Prediction*; Springer: New York, NY, USA, 2009.

-
33. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: London, UK, 2012.
 34. Ross, T.D.; Worrell, S.W.; Velten, V.J.; Mossing, J.C.; Bryant, M.L. Standard SAR ATR evaluation experiments using the MSTAR public release data set. In *Aerospace/Defense Sensing and Controls*; International Society for Optics and Photonics: Orlando, FL, USA, 1998; pp. 566–573.
 35. Gonzalez, R.C.; Woods, R.E.; Eddins, S.L. *Digital Image Processing Using MATLAB*; Pearson Education India: Noida, India, 2004.
 36. Maaten, L.V.D.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.