Multiview Deep Feature Learning Network for SAR Automatic Target Recognition

Jifang Pei [ID], Weibo Huo *, Chenwei Wang, Yulin Huang, Yin Zhang, Junjie Wu and Jianyu Yang

Department of Electrical Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China; jifangpei@uestc.edu.cn or peijfstudy@126.com (J.P.); 201811011927@std.uestc.edu.cn (C.W.); yulinhuang@uestc.edu.cn (Y.H.); yinzhang@uestc.edu.cn (Y.Z.); junjie_wu@uestc.edu.cn (J.W.); jyyang@uestc.edu.cn (J.Y.)
* Correspondence: hwbuyi@uestc.edu.cn or hwbuyi@163.com

Abstract: Multiview synthetic aperture radar (SAR) images contain much richer information for automatic target recognition (ATR) than a single-view one. It is desirable to establish a reasonable multiview ATR scheme and design effective ATR algorithm to thoroughly learn and extract that classification information, so that superior SAR ATR performance can be achieved. Hence, a general processing framework applicable for a multiview SAR ATR pattern is first given in this paper, which can provide an effective approach to ATR system design. Then, a new ATR method using a multiview deep feature learning network is designed based on the proposed multiview ATR framework. The proposed neural network is with a multiple input parallel topology and some distinct deep feature learning modules, with which significant classification features, the intra-view and inter-view features existing in the input multiview SAR images, will be learned simultaneously and thoroughly. Therefore, the proposed multiview deep feature learning network can achieve an excellent SAR ATR performance. Experimental results have shown the superiorities of the proposed multiview SAR ATR method under various operating conditions.

1. Introduction

Synthetic aperture radar (SAR) has been an important and powerful modern microwave sensor system in both military and civilian areas [1]. Due to its superior operational capabilities [2,3], SAR has played a significant information acquisition role for reconnaissance and detection nowadays. In addition, SAR can obtain the electromagnetic scattering characteristics of the detected targets and scenarios and acquire unique information from the imaging results at microwave frequencies [4], which have been of remarkable superiorities compared with other sensor systems.

With the improvement of the imaging capability of SAR systems, people have been interested in not only SAR signal processing but also interpretation or recognition of the real-world targets from SAR images. Automatic target recognition (ATR) [5–8] has become one of the most attractive but challenging research hotspots in SAR application. From the point of view of the users, an ideal ATR system should locate the regions with potential targets of interests from the SAR image and give those targets with accurate category labels intelligently and efficiently [9].

The general scheme of an end-to-end SAR ATR system, proposed by the researchers from MIT Lincoln Laboratory, has three basic stages with a hierarchical processing [10], i.e., detection [11], discrimination [12], and classification [13]. It aims to find the regions of interests (ROIs) from the SAR imagery, screen the targets we wanted [14], remove the false alarm clutters, and finally assign the classified attributes for the SAR targets with a well-designed classifier. In order to make the intelligent SAR target recognition to a reality,

people have proposed many novel SAR ATR methods in the past several years [15–26], such as principal component analysis (PCA) [27], linear discriminant analysis (LDA) [28], support vector machine (SVM) [15], adaptive boosting (AdaBoost) [29], conditional Gaussian model (CGM) [30], sparse representation [19], iterative graph thickening (IGT) [20], and so on, which generally performed well in applications.

To be viewed from a system level, there are generally two different SAR ATR genres according to the implementations: template-based [10] and model-based [31]. The template-based ATR system relies on template matching between the labeled target templates or features and the class-unknown input SAR target. It is a sequential processing and has advantages in simple construction and execution efficiency. Nevertheless, it is lacking in sufficient knowledge and intelligence, and its recognition results could be interfered by the operating condition variation and the template matching means. On the contrary, the model-based system includes two modules, i.e., offline model construction and online prediction and recognition, and tries to take a different approach to SAR ATR. It has much more intelligence and flexibility than the template-based system. However, additional complexity from the model construction and online prediction will also bring big challenges for the model-based SAR ATR system.

The ATR approaches mentioned above often need to extract specialized features from SAR images and predesign complex or sophisticated algorithms for target recognition. With the development of artificial intelligence, a novel ATR genre based on deep learning [32] has been growing fast and achieved remarkable performance in computer vision [33], natural language processing, and image classification [34] domains. As a new type of ATR approach, it can spontaneously discover and extract hierarchical and useful features from input data and give effective solutions to complex target recognition tasks. Naturally, due to the superiorities of deep learning, many significant works based on deep neural networks have also greatly upgraded the performance of SAR ATR [16,21,22,24,35].

Most of the existing SAR ATR systems and algorithms regard the SAR images as independent individuals, and they are often designed for single-input SAR images in practical ATR missions. Actually, it is a complex electromagnetic inverse scattering process for SAR imagery formation, and SAR images of the same target are often sensitive to different viewing angles. Hence, it is difficult for us to mine enough information from a single-input SAR image for ATR in general. On the other hand, SAR ATR will be benefited from multiview measurements [36] because the multiview SAR images of the same target could contain much richer classification information than single-view ones. SAR sensors have the abilities of obtaining images of the same target from different views with spotlight or circular modes in practice. Naturally, if the classification information could be effectively exploited or learned from the multiview SAR images, the SAR ATR performance may be significantly improved.

Inspired by this thought, a number of novel methods using multiview inputs have been proposed in recent years, which are of high recognition accuracy. For example, Ref. [36] shows the benefits of aspect diversity for SAR ATR based on the experimental analysis, and Ref. [37] proposes a multiview SAR ATR method using a Bayesian classifier, which improves the ATR performance. In Ref. [38], a machine learning based method is proposed for SAR ATR using multiple acquisition from multiple sensors, which improves the SAR target recognition performance greatly. Ref. [39] extracts the feature of the multiview SAR images using PCA and obtains a good classification result based on a radial basis function neural network. In [40], two fusion strategies are involved for target recognition with multiview SAR images, and the recognition performance excels the single-view based methods. Ref. [41] employs joint sparse representation and proposes a novel multiview SAR ATR method, and the experimental result shows its superiority. Ref. [42] exploits the multiview SAR robust target recognition and further improves the ATR performance based on sparse representation classification. Some multiview SAR ATR methods are also proposed based on various deep neural network architectures [43–45], which could achieve outstanding recognition results under different operating conditions.

Generally, multiview SAR ATR is a complex and integrated information processing procedure. In order to achieve outstanding multiview SAR ATR performance, two important issues must be incorporated: a valid ATR processing framework and an appropriate ATR algorithm for classification feature learning from limited raw SAR samples. A reasonable processing framework is necessary for the effectiveness of multiview SAR ATR, while the ATR algorithm is one of the most key points in the framework. Hence, it is indispensable and desirable to establish a standard processing framework for multiview SAR ATR architecture design and then search for an effective ATR algorithm.

In this paper, we will give a general processing framework for multiview SAR ATR including three parts, i.e., raw multiview SAR data formation, multiview SAR data preprocessing, and multiview target recognition, which can provide an effective and standard way to multiview SAR ATR system design. Then, a novel ATR method using a multiview deep feature learning network is proposed based on this framework. The proposed deep neural network is with a multiple input parallel topology, and some specific modules such as convolutional layer, convolutional gated recurrent unit (ConvGRU), weighted concatenation unit (WCU), 3D convolutional layer, and 3D pooling layer are embedded in this network. Both the intra-view and inter-view features of the input multiview SAR images will be thoroughly learned with this elaborately designed multiview deep feature learning network. Therefore, the proposed network can take advantage of comprehensive and significant classification information from multiview SAR images and achieve high target recognition accuracy.

The main contributions compared with available SAR ATR works are the following: (1) We give a general processing framework for multiview SAR ATR, which can make a paradigm for ATR system designs and future studies of this field. (2) A multiview deep feature learning network is proposed for effective SAR ATR, and this network can simultaneously extract the intra-view and inter-view features from multiview SAR images. (3) Compared with the available SAR ATR methods, the proposed deep neural network can achieve excellent ATR performances under various operating conditions but with limited raw SAR data for training sample generation.

This paper is organized as follows: A general processing framework for multiview SAR ATR is introduced in Section 2. Section 3 details the proposed SAR ATR method using a multiview deep feature learning network. Experiments are carried out in Section 4, and Section 5 gives the conclusions of our work.

2. Multiview SAR ATR Processing Framework

Practical implementation of SAR ATR had been summarized as a multistage processing by the researchers from MIT Lincoln Laboratory in the last century [46], which is a classical and excellent SAR ATR framework. Nevertheless, that ATR scheme was generalized and mainly designed for single-view input SAR image at the beginning. Multiview SAR data are with higher dimensions than single-view ones, and contain rich classification information, so this needs a more sophisticated and specific processing ATR scheme than before. Therefore, based on the MIT ATR scheme, we give a general processing framework that is appropriate for multiview SAR ATR.

The framework includes three specific parts as shown in Figure 1, i.e., raw multiview SAR data formation, multiview SAR data preprocessing, and multiview target recognition, each of which performs easily identifiable functions. Modules in this framework are detailed as follows.
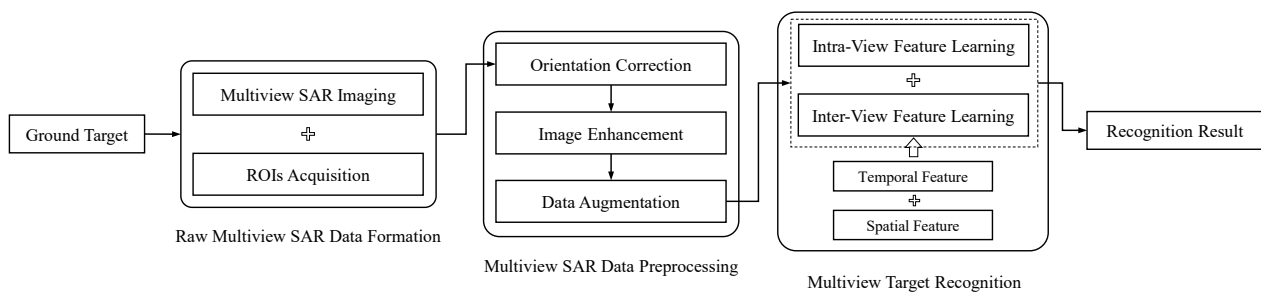
Figure 1. Basic scheme of multiview SAR ATR processing framework.

2.1. Raw Multiview SAR Data Formation

The first module in the framework is to acquire the eligible and valid raw multiview SAR images and find out the ROIs, which can locate the targets we wanted and reduce the computational load of the ATR system. Generally, this module should contain two processing steps, i.e., multiview SAR imaging and ROIs acquisition. Some SAR imaging modes, such as spotlight mode [47] and circular mode [48], can continuously observe the same scene or target and are perfect for raw multiview SAR images collection. Then, the target chips with multiple views will be obtained by the ROI acquisition step, and there are many target detection and discrimination methods that can be chosen to realize it.

2.2. Multiview SAR Data Preprocessing

After the raw multiview SAR data formation, the multiview SAR target chips are obtained; however, there are still some problems to be solved. For example, the orientations of the same target on the SAR chips are different, and the scattering information of the target on the multiview SAR images could be inapparent. In addition, sufficient training samples should be fed into the multiview SAR ATR algorithm to optimize its parameters during the training phase. However, the amount of the available raw multiview SAR data are often limited in practice, which could lead to overfitting of the ATR algorithm.

The aims of the multiview SAR data preprocessing are to eliminate the inconsistence, enhance the scattering information, and augment the raw multiview SAR data for training, which correspond to orientation correction, image enhancement, and data augmentation in this module, respectively. After the data preprocessing, the multiview SAR data are more suitable for the following ATR processing, and the classification information of the multiview SAR targets will be more easy to learn than before.

2.3. Multiview Target Recognition

Multiview target recognition is the back-end module in the multiview SAR ATR processing framework. It constructs ATR algorithms, receives the multiview SAR samples from the preceding module, and assigns the most probable classified label for the target. Essentially, this module is to learn and extract effective classification features from the input samples and make optimal division for the features with hyperplanes in the feature space.

There are two kinds of very important features to be learned in multiview SAR images, i.e., intra-view feature and inter-view feature. The intra-view feature means the inherent scattering or structural feature of the SAR target within each view, while the inter-view feature is the mutual feature in the multiview SAR image sequence, which is distinct from single-view SAR ATR. Meanwhile, the inter-view feature includes two individual features. When SAR observes the same target from different views, the correlated feature among the multiview image sequence, namely the temporal feature, will contain intrinsic classification information. In addition, the variation feature of the multiview image sequence, i.e., spatial feature, can also provide complementary discriminative information of the same target and benefit to ATR. Therefore, the most important point in multiview target recognition module is to design an appropriate ATR algorithm to simultaneously learn classification features of

both intra-view and inter-view from multiview SAR images. After feature learning, the multiview target recognition module will give us an accurate class attribute of the target.

Thus far, the multiview SAR ATR processing framework is summarized as three individual but related modules with several distinct steps. In this way, the multiview SAR ATR problem can be effectively handled. While this ATR framework includes some specific processing steps within each module, it is noted that not every processing step is absolutely necessary; people could also make some adjustments in ATR practice.

3. Proposed Multiview SAR ATR Method

A multiview deep feature learning network is presented for SAR ATR in this section, which is based on the above-mentioned framework and can simultaneously learn both the intra-view and inter-view features of multiview SAR images. According to the multiview ATR framework, we will first discuss the raw multiview SAR data formation and multiview SAR data preprocessing. Then, the design of the feature learning network for multiview SAR ATR will be given, and at last the configuration of the network will also be detailed.

3.1. Raw Data Formation

In the multiview SAR imaging pattern, the SAR sensor collects returns and obtains the multiview images for a given target from different elevation and aspect angles. For simplicity, the depression angle is set as constant here. Figure 2 shows the geometric model of the multiview SAR imaging process. The given view interval is denoted as θ, and the view number is $k > 1$. Then, the target chips with multiple views will be obtained by target detection and discrimination methods. Using these raw SAR images from different view angles, more classification information could be exploited than from the single-view pattern.
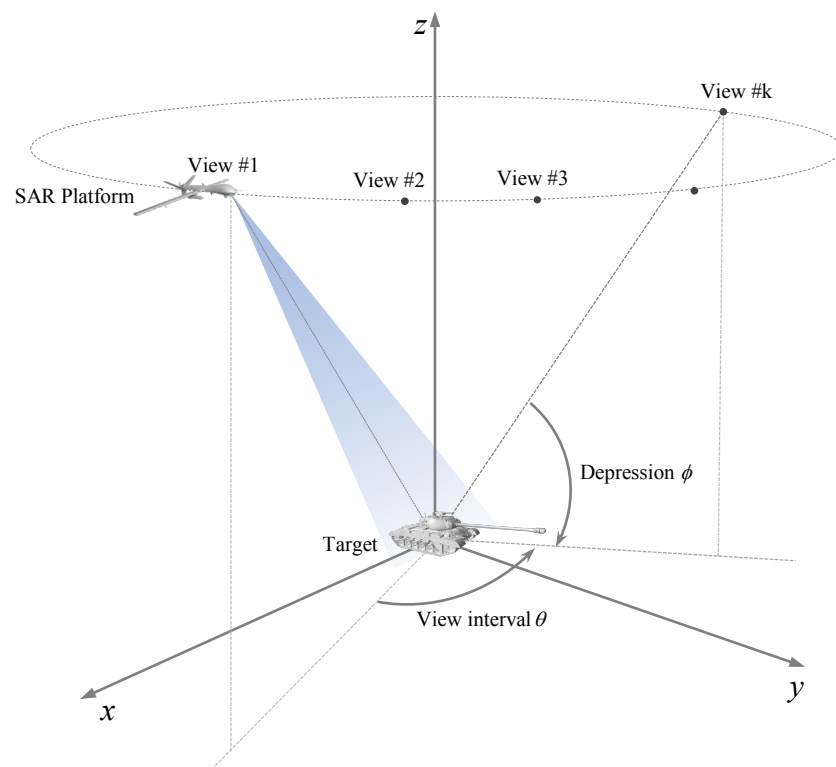

Figure 2. Geometric model of multiview SAR ATR for a ground target.

3.2. Data Preprocessing

Before the training and recognition phases of ATR, some preprocessing steps are needed for the raw multiview SAR data, which includes orientation correction, image

enhancement, and data augmentation. The targets on SAR images are often sensitive to the views and present themselves differently both in scattering characteristics and orientations on the multiview images. In order to keep the scatting information of the targets from multiview images while reducing their orientation difference, we align the targets to the same orientation by aspect rotation with an affine transformation:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} \tag{1}$$

where ϕ is the rotation angle estimated by the SAR target aspect estimation method [49], (p, q) is the original coordinates, and (u, v) is the transformed coordinates.

After target orientation correction, we employ the gray enhancement method based on power function [50] to enhance the scattering information of the raw multiview SAR images:

$$\mathbf{x}(u, v) = \left[\mathbf{x}'(u, v)\right]^{\rho} \tag{2}$$

where $\mathbf{x}'(u, v)$ is the original image and $\mathbf{x}(u, v)$ is the SAR image after information enhancement, respectively, and ρ is the enhancement factor.

Next, we use a typical multiview SAR data augmentation method to generate adequate samples for training [45]. Suppose $\mathbf{X}^{(raw)} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_C\}$ is the set of raw SAR images. The image set $\mathbf{X}_i = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_i}\}$ collected by SAR with their aspect angles $0°-360°$ belongs to class y_i. Their corresponding aspect angles are $\varphi(\mathbf{x}_i)$. The class label set is $\{y_i \in [1, 2, \cdots, C]\}$. For a given view number $k > 1$, all the view combinations of one class SAR images can be obtained, and the combination number is $C_{n_i}^k = n_i!/k!(n_i - k)!$. Then, the training data can be significantly augmented as the steps in Table 1. In Table 1, N_i is the number of the multiview SAR image sequence from the combinations that fulfills the selection condition in class y_i, and the final input multiview SAR image sequence set for training is $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^C\}$.

Table 1. Process of multiview SAR data augmentation.

Initialization: view interval θ and view number k.

Input: raw SAR images $\mathbf{X}^{(raw)} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_C\}$ and class labels $\{y_i \in [1, 2, \cdots, C]\}$.

for $i = 1$ to C do

    Find out all view combinations of raw SAR images from class y_i.

    for $j = 1$ to $C_{n_i}^k$ do

        Arrange elements for each combination $\left\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \cdots, \mathbf{x}_{j_k}\right\}$,

        $s.t. \left(\varphi\left(\mathbf{x}_{j_1}\right) < \varphi\left(\mathbf{x}_{j_2}\right) < \cdots < \varphi\left(\mathbf{x}_{j_k}\right) \text{ or } \varphi\left(\mathbf{x}_{j_1}\right) > \varphi\left(\mathbf{x}_{j_2}\right) > \cdots > \varphi\left(\mathbf{x}_{j_k}\right)\right).$

        Find out $\mathbf{X}_j^i$, $s.t. \left|\varphi\left(\mathbf{x}_{j_1}\right) - \varphi\left(\mathbf{x}_{j_k}\right)\right| \leq \theta.$

    end for

  Get $\mathbf{X}^i = \left\{\mathbf{X}_1^i, \mathbf{X}_2^i, \cdots, \mathbf{X}_{N_i}^i\right\}.$

end for

Output: multi-view SAR image sequence set $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^C\}$ for ATR algorithm training.

Figure 3 shows an example of using the data augmentation method to generate multiview SAR samples for training. Suppose the view number is 3 in each view interval, the data augmentation method can get nine 3-view SAR training samples from only six raw SAR images. When the view number and interval increase, sufficient training samples can be obtained for a given number of raw SAR images.
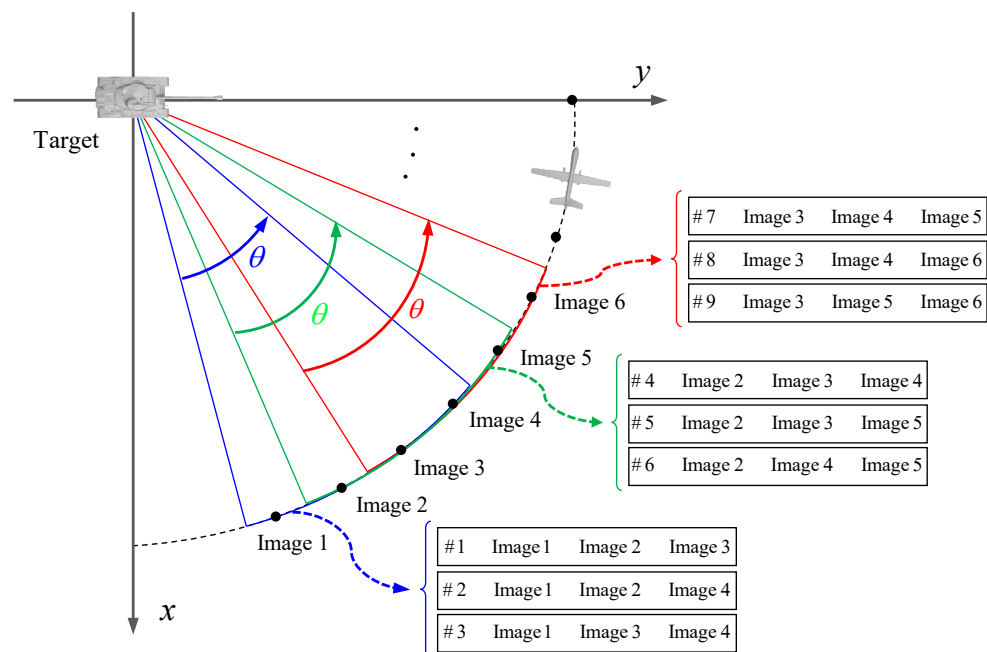
Figure 3. Graphic explanation of the data augmentation method.

3.3. Multiview Deep Feature Learning Network

The basic architecture of the proposed multiview deep feature learning network is shown in Figure 4. As we can see, this deep neural network employs a parallel topological structure with multiview inputs, and these multiple inputs are progressively merged and fused in different layers, which can effectively learn and extract the recognition information from the SAR images with different views.
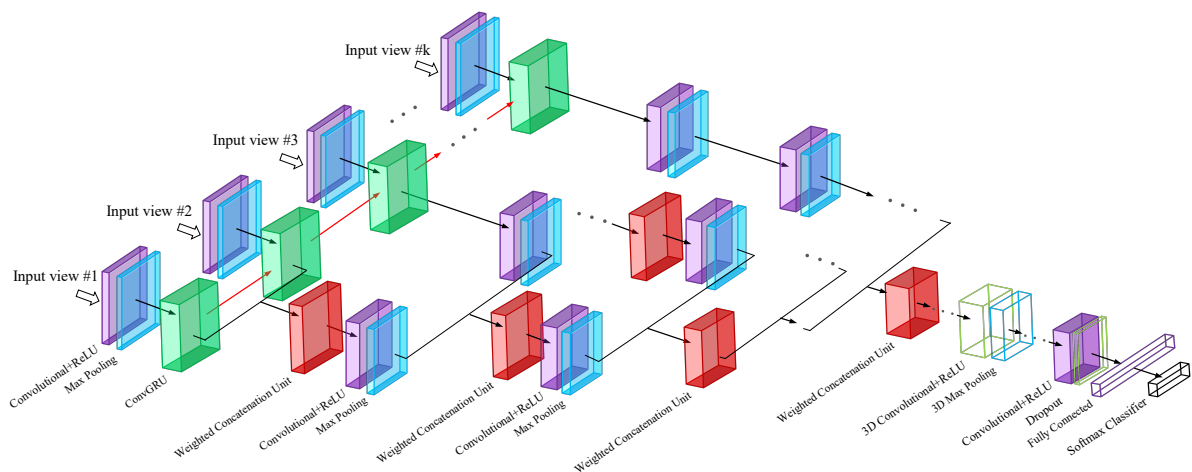
Figure 4. Basic architecture of the proposed multiview SAR ATR network.

As is mentioned in Section 2.3, intra-view and inter-view features are two kinds of important features that should be learned in multiview SAR images, which incorporate complete classification information. The proposed multiview deep feature learning network is designed with alternate convolutional and pooling layers in each branch, which can learn the inherent classification feature of the target in the SAR image within each view and reduce the feature dimension. Meanwhile, a special recurrent neural network structure, ConvGRU, is implanted in the proposed deep neural network to extract the temporal feature and learn the correlated feature among the multiview SAR images. In addition, we propose a new spatial feature learning module, WCU, to effectively learn and fuse the classification feature of the multiview SAR images between two different network

branches. After all the network branches being merged, the feature maps are fed into a 3D convolutional layer followed by a 3D max pooling layer to further learn the inter-view feature from the fused feature maps. Finally, the proposed network ends with a fully connected layer, and the recognition decision is conducted by the softmax classifier.

From the basic architecture, it can be seen that the proposed multiview deep feature learning network is capable of learning both intra-view and inter-view classification features from multiview SAR images and benefits ATR. In the following discussions, we will detail the network layers for intra-view feature and inter-view feature learning in the multiview deep feature learning network.

The layers designed for intra-view feature learning from multiview SAR images mainly include convolutional layer and pooling layer which are detailed as follows.

3.3.1. Convolutional Layer

Convolutional layer is inspired by the process of the biological neuron in the visual cortex to a specific stimulus and can learn the features from images well. Thus, it can be used to effectively learn the intra-view feature from the multiview SAR images. Here, let $\mathbf{a}_i^{(l-1)}$ be the ith feature map in $l-1$ convolutional layer in our neural network, and it will be connected to all the feature maps in the l layer by convolution operation. $\mathbf{w}_{ij}^{(l)}$ denotes the convolution kernel operating the ith input feature map to the jth output feature map in the l convolutional layer, and $b_j^{(l)}$ is the bias. The forward process in the convolutional layer for each unit can be written as

$$\mathbf{z}_j^{(l)} = \sum_i \mathbf{a}_i^{(l-1)} * \mathbf{w}_{ij}^{(l)} + b_j^{(l)} \tag{3}$$

where $\mathbf{z}_j^{(l)}$ is the output feature map, the symbol $*$ is 2D convolution operation. Rectified linear units (ReLUs) are selected as the activation function after each convolutional layer, which can increase the nonlinear properties of the proposed network. The ReLU can be expressed as

$$\mathbf{a}_j^{(l)} = \sigma\left(\mathbf{z}_j^{(l)}\right) = \max\left\{0, \mathbf{z}_j^{(l)}\right\} \tag{4}$$

3.3.2. Pooling Layer

Convolutional layer is often followed by a pooling layer in the structure of a deep neural network. The pooling layer can select the local feature from its input feature map and reduce the dimension of the feature, which is a perfect auxiliary for intra-view feature learning. Here, we utilize max pooling operation in the proposed neural network, which can be written as

$$\mathbf{a}_j^{(l)}(x,y) = \max\left[\mathbf{a}_j^{(l)}(x \cdot s_1 + u, y \cdot s_2 + v)\right]\big|_{0 \le u < p_1, 0 \le v < p_2} \tag{5}$$

where s_1 and s_2 are the pooling strides, and p_1 and p_2 are the pooling window sizes. When the pooling window on the feature map slides and computes the maximum in the window as its output, the valuable local feature of the SAR images within each view is effectively extracted, and the feature dimension is also reduced.

As the intra-view feature of the multiview SAR images is being learned by the convolutional layer and max pooling layer, the modules for inter-view feature learning are also designed in the proposed neural network, which are described in the following.

3.3.3. Convolutional Gated Recurrent Unit

Multiview SAR images contain the temporal feature among the sequence, which could provide intrinsic classification information. Thus, the ATR performance will be improved if we can extract it well from multiview SAR images. A recurrent neural network is a special deep neural network that is suitable to learn the correlated feature of data

sequence, which can handle this problem. A typical recurrent neural network structure, gated recurrent unit (GRU), is able to adaptively extract the temporal feature by resetting and updating the flow of information inside the unit, which has the potential of capture dependencies among the multiview SAR image sequence. However, classical GRU is with a fully-connected operation within each unit and cannot take advantage of the underlying structural information of the feature maps learned from SAR images. We therefore employ a kind of GRU with convolution operation, namely ConvGRU [51], to the proposed deep neural network. It can obtain the superiority of recurrent unit but is more appropriate for feature learning from multiview SAR images than classical ones.

The block diagram of ConvGRU is a two-input and two-output system and mainly composed of reset gate, update gate, and some other operations, which is shown in Figure 5. In the ConvGRU, when a new input arrives, the reset gate will control the feature learned from a previous view we might want to remember, while the update gate will determine how much the new feature of current view will be retained.
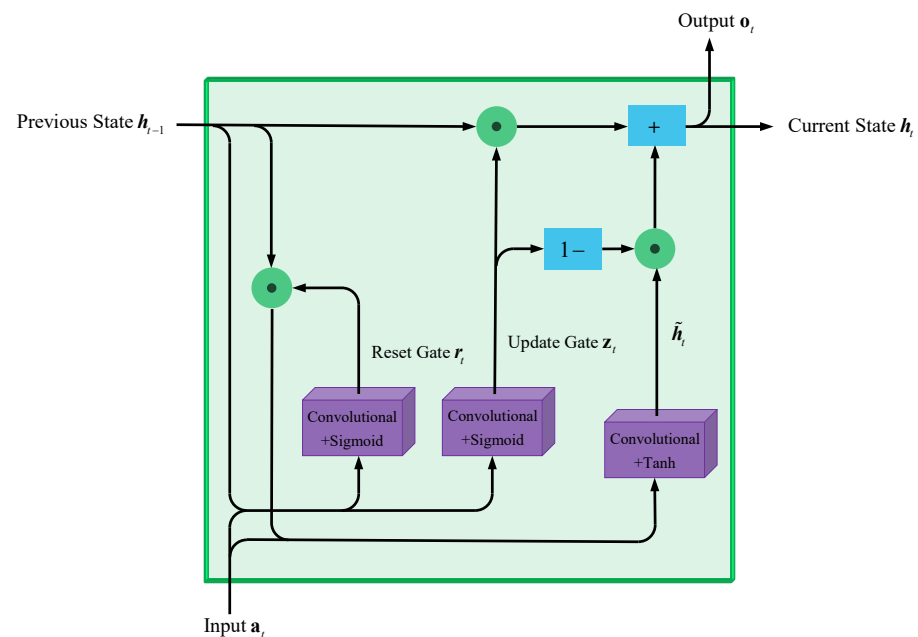

Figure 5. Basic block diagram of ConvGRU.

For a given ConvGRU, the input feature map is $\mathbf{a}_t$ and the state input of the previous learned feature is $\mathbf{h}_{t-1}$. The reset gate $\mathbf{r}_t$ and update gate $\mathbf{z}_t$ are computed as

$$\mathbf{r}_t = sig(\mathbf{w}_{ar} * \mathbf{a}_t + \mathbf{w}_{hr} * \mathbf{h}_{t-1} + \mathbf{b}_r) \tag{6}$$

$$\mathbf{z}_t = sig(\mathbf{w}_{az} * \mathbf{a}_t + \mathbf{w}_{hz} * \mathbf{h}_{t-1} + \mathbf{b}_z) \tag{7}$$

where $\mathbf{w}_{ar}$, $\mathbf{w}_{hr}$ and $\mathbf{w}_{az}$, $\mathbf{w}_{hz}$ are the convolution kernel, and $\mathbf{b}_r$, $\mathbf{b}_z$ are their corresponding biases, respectively. $sig(\cdot)$ denotes a sigmoid function to transform input values to the interval $(0,1)$. Then, the candidate hidden state of the ConvGRU can be computed as

$$\tilde{\mathbf{h}}_t = tanh[\mathbf{w}_{ah} * \mathbf{a}_t + \mathbf{w}_{hh} * (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h] \tag{8}$$

where $\mathbf{w}_{ah}$ and $\mathbf{w}_{hh}$ are the convolution kernel, and $\mathbf{b}_h$ is the bias. The symbol $\odot$ indicates Hadamard product, and $tanh(\cdot)$ denotes a tanh function to ensure the values of the candidate hidden state to the interval $(-1,1)$.

Finally, the current state $\mathbf{h}_t$ and the output of the ConvGRU are obtained by the following equation:

$$\mathbf{o}_t = \mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t \tag{9}$$

When the feature maps from previous network module arrive, the temporal feature will be effectively learned from the multiview SAR image sequence by the ConvGRUs.

3.3.4. Weighted Concatenation Unit

The inputs or the feature maps are progressively merged and fused in different layers in the proposed neural network. Thus, several concatenating operations should exist in the network to extract classification information from different views. The input feature maps are straightforwardly stacked in a traditional concatenation module, and the importance of each feature map is treated as equal. However, the features learned from multiview SAR images are different and their corresponding feature maps are of variation in different network branches. Therefore, it is necessary to select beneficial spatial features, focus on important features, and suppress trivial ones from different network branches during feature learning. To this end, a new spatial feature learning module, WCU, is designed in the neural network to learn and fuse classification feature from different network branches. The block diagram of the proposed WCU is shown in Figure 6.
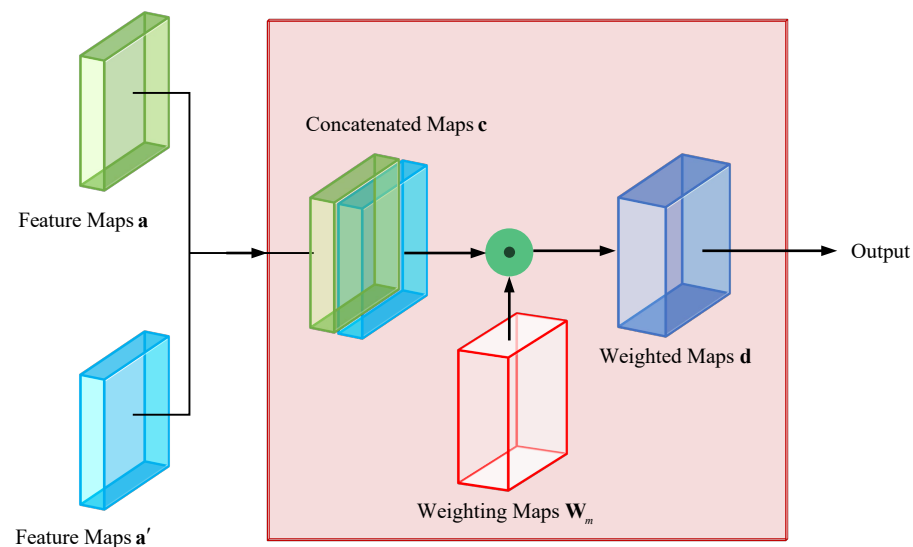
Figure 6. Basic block diagram of WCU.

Let feature maps $\mathbf{a} \in R^{m \times n \times r}$ and $\mathbf{a}' \in R^{m \times n \times s}$ be two inputs flowed into the WCU, and the feedforward propagation in the WCU is processed as

$$\begin{cases} \mathbf{c} = conc(\mathbf{a}, \mathbf{a}') \\ \mathbf{d} = \mathbf{W}_m \odot \mathbf{c} \end{cases} \tag{10}$$

where $\mathbf{c} \in R^{m \times n \times (r+s)}$, function $conc(\cdot, \cdot)$ denotes concatenation operation as in Figure 6. $\mathbf{W}_m \in R^{m \times n \times (r+s)}$ is the corresponding weighting maps to be learned, and the feature maps $\mathbf{d} \in R^{m \times n \times (r+s)}$ are the output weighted maps of the WCU.

Through the weighted concatenation processing, WCU is able to find out the optimal weighting maps during network learning and concatenate and weight the input feature maps, which can learn and emphasize meaningful spatial features of multiview SAR images.

3.3.5. 3D Convolutional Layer and 3D Pooling Layer

After all the input views have been progressively merged in the network, the learned feature maps are concatenated together. Then, those concatenated feature maps will flow into a 3D convolutional layer [52] and a 3D max pooling layer, in which the inter-view

feature from the fused feature maps will be further learned. In contrast to 2D convolution, the input, convolution kernel, and the output are represented as 3D tensors in the 3D convolutional layer, which can learn and fuse both the spatial and temporal features from the preceding concatenated feature maps.

Specifically, the convolution is in the form of 3D calculation on feature tensors in 3D convolutional layer, which can be written as

$$\mathbf{a}_j^{(l)}(x,y,z) = \sigma \left(\sum_i \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} \mathbf{a}_i^{(l-1)}(x-p,y-q,z-r) \mathbf{w}_{ij}^{(l)}(p,q,r) + b_j^{(l)} \right) \qquad (11)$$

where $\mathbf{a}_j^{(l)}(x,y,z)$ denotes the value of the output feature tensor at position (x,y,z), $\mathbf{w}_{ij}^{(l)}$ here is the 3D convolution kernel with a size of $P \times Q \times R$, and $\sigma(\cdot)$ is the ReLU nonlinear activation function.

Usually, a 3D convolutional layer is followed by a 3D pooling operation to extract the local feature and reduce the dimension of the feature tensor. Here, the 3D max pooling operation is employed after the 3D convolutional layer, which can be expressed as

$$\mathbf{a}_j^{(l)}(x,y,z) = \max \left[\mathbf{a}_j^{(l)}(x \cdot s_1 + u, y \cdot s_2 + v, z \cdot s_3 + w) \right] \qquad (12)$$

where s_1, s_2 and s_3 are the pooling strides, $0 \leq u < p_1$, $0 \leq v < p_2$, $0 \leq w < p_3$, and p_1, p_2 and p_3 are the pooling window sizes.

In addition, some other helpful module or operation, such as dropout and softmax classifier, are necessary in the proposed network. Dropout operation [53] is a good choice to reduce overfitting and widely used in neural network. It forces the network neurons to have robust learning ability with the random active neuron combinations. In our deep neural network, dropout operation is included after the last convolutional layer to increase the generalization.

After all the intra-view and inter-view features of the multiview SAR images have been learned, those feature maps are transformed as a feature vector connecting to a fully connected layer. Then, the softmax classifier is used for the final recognition:

$$p\left(y_i | \mathbf{z}^{(L)}\right) = \frac{\exp\left(z_i^{(L)}\right)}{\sum_{j=1}^{C} \exp\left(z_j^{(L)}\right)} \qquad (13)$$

where $\mathbf{z}^{(L)}$ is the input feature vector to the softmax classifier, and C is the class number. Finally, the recognition result corresponds to the class with the maximal posterior probability.

3.3.6. Cost Function and Network Training

After the forward propagation, the proposed deep network will compare the class label with the inferred output of the softmax classifier, which is calculated with the cross entropy cost function:

$$L = - \sum_{i=1}^{C} y_i \log p\left(y_i | \mathbf{z}^{(L)}\right) \qquad (14)$$

The training method to minimize the cost function and optimize those trainable parameters is similar to the common SAR ATR neural networks although the proposed deep neural network has a complex network structure. The back propagation through time algorithm [54] can be used to train the network parameters for temporal feature learning module, i.e., ConvGRU, while training the rest of the proposed deep network is realized by a back propagation algorithm. Once the network training phase finished, the proposed deep network will get its optimal parameters, which can effectively learn various features and make accurate classification for the input multiview SAR images.

4. Experiments and Results

ATR performance of the proposed multiview deep feature learning network will be evaluated in this section. First, the network architecture setup is specified, and the multiview SAR training and testing data formation for experiments are also given. Finally, we will extensively assess the performance of the proposed multiview deep network under different SAR ATR operating conditions.

4.1. Network Architecture Setup

We will utilize two network instances with three and four input views to assess the proposed deep network for SAR ATR, which are shown in Figure 7. The input SAR image size for a three view network instance is 90×90 and for four view network instance is 120×120. The stride size in each convolutional layer is 1×1, and in each pooling layer is set as 2×2. The probability of dropout is 0.5 during training phase. Other hyper-parameters in the proposed multiview deep network are shown in Figure 7, and those hyper-parameters in the instance are determined by statistical validation and trials.

The proposed multiview neural network is implemented with the framework of TensorFlow. The instances are trained with a minibatch size of 128 examples and learning rate 0.001. Their weights and biases are initialized from Gaussian distributions with zero mean and a standard deviation of 0.1. All the experiments are conducted on a PC with Intel Core i9-10900K CPU at 3.70 GHz, 64.0G RAM, and a NVIDIA GeForce RTX 3080 GPU.
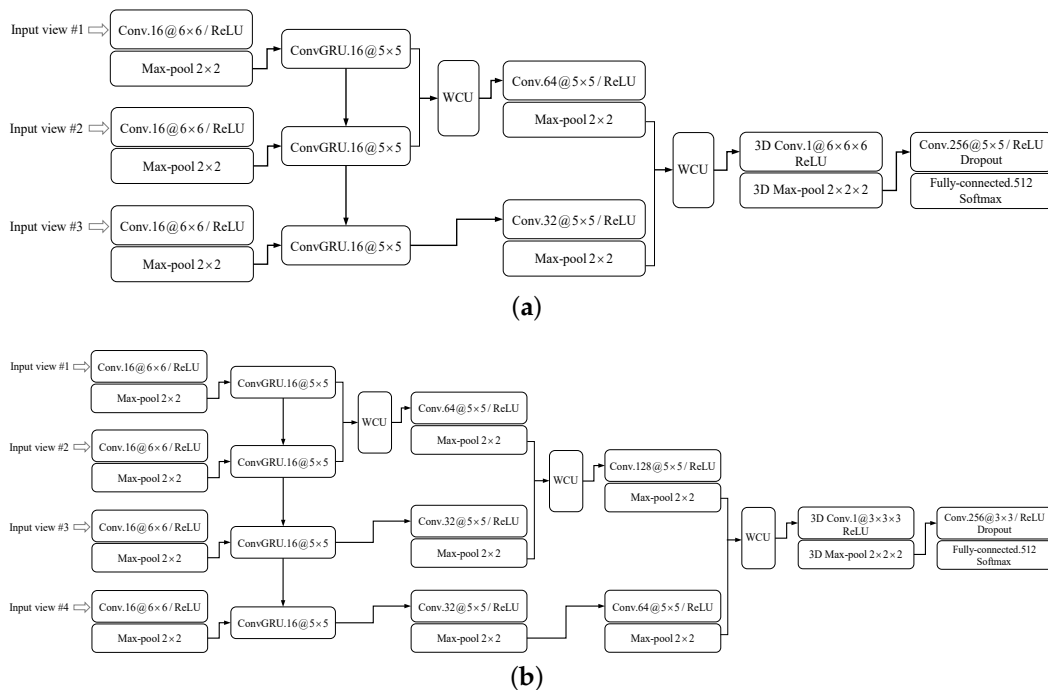
Figure 7. Network instances with three and four input views. A convolutional layer is expressed as Conv. (number of feature maps)@(convolution kernel size), and Max-pool $p_1 \times p_2$ represents its pooling window size. (**a**) network instance with three input views; (**b**) network instance with four input views.

4.2. Data Set

In our experiment, raw SAR images from the moving and stationary target acquisition and recognition (MSTAR) program are employed to assess the recognition performance of the proposed multiview deep neural network. The MSTAR program aims to develop the advanced SAR ATR system in a battlefield environment under the support of U.S. Defense Advanced Research Projects Agency and the U.S. Air Force Research Laboratory [55]. It has collected a significant quantity of SAR images as the benchmark data set to evaluate the performance of an advanced SAR ATR system, and those SAR images were acquired

near Huntsville, AL, USA, by the Sandia National Laboratory using the Synthetic Aperture Radar Target Location and Recognition System. The MSTAR data set includes a series of 0.3 m × 0.3 m resolution SAR images collected with an X-band spotlight SAR sensor. Those images contain different types of vehicle targets and clutter, and ten classes of targets, including T62 and T72 tanks, 2S1 rocket launcher, ZIL131 truck, BTR70, BTR60, BRDM2 and BMP2 armored personnel carriers, ZSU23/4 air defense unit, and D7 bulldozer, are utilized in our experiment for ATR performance evaluation. The optical images of those targets and their corresponding SAR images are illustrated in Figure 8.

The proposed multiview deep feature learning network will be tested both under standard operating condition (SOC) and extended operating condition (EOC) to comprehensively evaluate its recognition performance. In the following Sections 4.3 and 4.4, the two instances of the proposed deep network will be tested under SOC and EOC, respectively. In addition, we will also compare the recognition performance of the proposed deep network with some new published and widely cited SAR ATR algorithms in Section 4.5.

Figure 8. Optical images and MSTAR SAR images of targets. (**a**,**b**) optical images and their corresponding SAR images for BMP2, BTR70, T72, BTR60, 2S1; (**c**,**d**) optical images and their corresponding SAR images for BRDM2, D7, T62, ZIL131, ZSU23/4.

4.3. Results under SOC

In this experiment, we will evaluate the recognition performance of the network instances with ten classes of typical vehicle targets under SOC. We only select part of the raw SAR images with depression $17°$ from the MSTAR data set to generate multiview SAR image sequences for network training. Their aspect angles of those selected raw SAR images for each target type are all covered from $0°-360°$. All of the raw SAR images with depression $15°$ from the data set are used to generate testing samples. The usage of raw SAR images in this experiment for training and testing samples generation is listed in Table 2.

Table 2. Raw SAR images selection from the MSTAR dataset in experiments.

Training			Testing	
Target Types	**3-Views Instance**	**4-Views Instance**	**Target Types**	**3-/4-Views Instances**
BMP2sn-9563	78	59	BMP2sn-9563	195
BTR70	78	59	BTR70	196
T72sn-132	78	58	T72sn-132	196
BTR60	86	64	BTR60	195
2S1	100	75	2S1	274
BRDM2	100	75	BRDM2	274
D7	100	75	D7	274
T62	100	75	T62	273
ZIL131	100	75	ZIL131	274
ZSU23/4	100	75	ZSU23/4	274

Here, we use the method described in Section 3.2 to generate a large number of multiview SAR image sequences from a few subsets of the MSTAR data set for deep network training. The view interval θ is 45° in both the multiview training and testing phase. There are 48,764 and 43,533 multiview SAR image sequences with 17° depression for three and four input view deep network instances training, respectively. We randomly select the samples from multiview SAR image sequences with 15° depression for testing. For each class, the number of randomly selected testing samples is 2000, thus there are 20,000 tests for ATR performance evaluation in each SOC experiment.

The recognition results of the proposed deep networks with three and four input views are shown in Tables 3 and 4, which are presented with confusion matrices. The rows in confusion matrix represent the ground truths of the target labels, and its columns are the predicted class labels by the ATR method.

Table 3. Confusion Matrix of 3-views network instance under SOC (Recognition rate: 99.30%).

Class	BMP2sn-9563	BTR70	T72sn-132	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU23/4
BMP2sn-9563	99.60	0.10	0.00	0.10	0.10	0.00	0.00	0.10	0.00	0.00
BTR70	0.10	99.40	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
T72sn-132	0.20	0.00	99.40	0.00	0.00	0.00	0.00	0.40	0.00	0.00
BTR60	0.20	0.40	0.00	96.10	0.50	1.20	0.00	1.10	0.50	0.00
2S1	0.00	0.00	0.10	0.00	99.90	0.00	0.00	0.00	0.00	0.00
BRDM2	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
D7	0.00	0.00	0.00	0.00	0.00	0.00	99.70	0.00	0.30	0.00
T62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	99.60	0.40	0.00
ZIL131	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00
ZSU23/4	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.50	0.00	99.30

Table 4. Confusion Matrix of 4-views network instance under SOC (Recognition rate: 99.62%).

Class	BMP2sn-9563	BTR70	T72sn-132	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU23/4
BMP2sn-9563	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BTR70	0.30	99.50	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00
T72sn-132	0.10	0.00	99.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BTR60	0.40	0.40	0.00	98.70	0.20	0.30	0.00	0.00	0.00	0.00
2S1	0.20	0.00	0.00	0.00	99.80	0.00	0.00	0.00	0.00	0.00
BRDM2	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
D7	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00
T62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
ZIL131	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.70	98.30	0.00
ZSU23/4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

It can be observed that the recognition rates of the proposed multiview deep feature learning networks with three and four views are all higher than 99.00% under SOC in the ten classes problem. From the experimental results in Tables 3 and 4, we can see that the multiview SAR images are with much classification information, and the proposed multiview deep network is able to learn both the intra-view and inter-view classification features of these multiview SAR images but only with a few raw SAR data for training samples generation. Hence, we can come to a conclusion that the designed deep network architecture based on the general processing framework can obtain satisfactory recognition performances in the SOC ATR experiment.

Part of the input testing multiview SAR samples and their corresponding output tensors in the last fully connected layer are mapped into 2D Euclidean space by the t-distributed stochastic neighbor embedding (t-SNE) [56] algorithm to illustrate the good classification performance of the multiview deep feature learning network. T-SNE is a powerful dimension reduction algorithm, which can help us study the distribution characteristics of the high-dimensional data in a visualized low-dimensional space.

Figure 9 shows the results of 2D visualization example of the input testing multiview SAR samples and their corresponding outputs in the last fully connected layer of the two proposed networks. In Figure 9, the sample points with the same color belong to the same target class. Figure 9a,b illustrate the input multiview SAR samples and the corresponding output for a 3-view network instance, and Figure 9c,d show the input and output of the multiview SAR samples for the 4-view network instance, respectively.

From Figure 9a,c, we can see that the visualization results of the original multiview SAR samples are mixed together and difficult to be classified in practice. Nevertheless, after being processed by the proposed network, both the intra-view and inter-view classification features are learned, and the samples with the same class label get close, and the samples from different classes separate from each other in the visualized low-dimensional space, which makes them easy to be distinguished and leads to satisfactory recognition results.
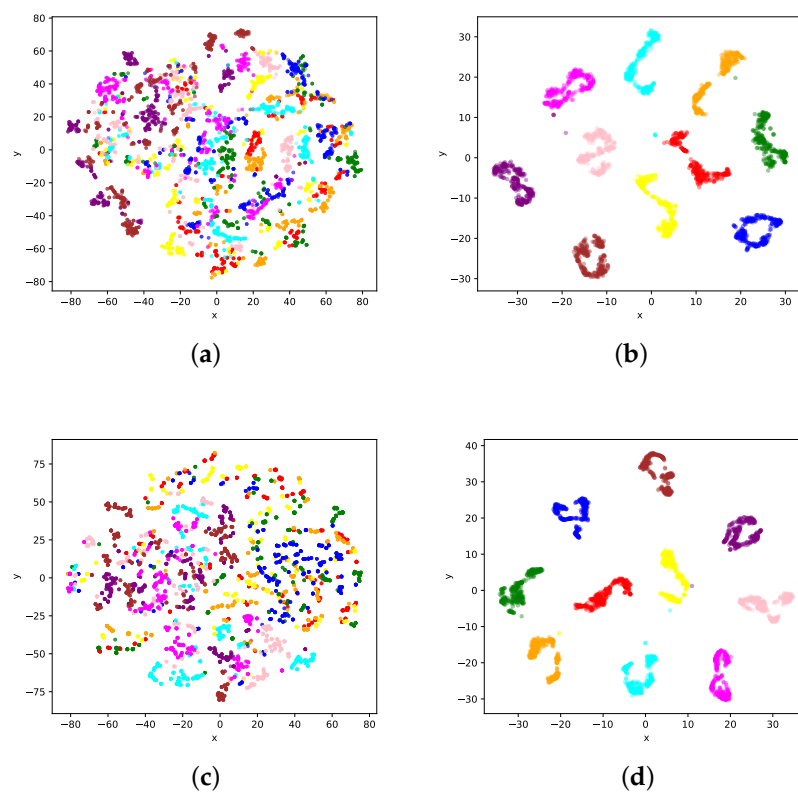

(a)

(b)

(c)

(d)

Figure 9. Visualization of input multiview SAR samples and their corresponding outputs in the last layer of proposed networks. Sample points with the same color belong to the same target class. (**a**,**b**) Input multiview SAR samples and corresponding output for 3-view network instance; (**c**,**d**) input multiview SAR samples and corresponding output for 4-view network instance.

4.4. Results under EOC

SAR ATR performances will be influenced by many kinds of operating condition variations in reality. Thus, we will do some experiments to assess the ATR performances of the multiview deep network with complex test scenarios under EOC. Here, we first evaluate the performance of the proposed deep network with the testing data with large depression angle variation denoted as EOC-D. The selection of raw SAR images in this experiment is listed in Table 5. In this test, four types of ground targets, 2S1, BRDM-2, T-72sn-132 and ZSU-234, in the MSTAR data set with depression 17° are selected for training samples generation. Thus, there are 20,951 and 19,075 multiview SAR image sequences for three and four input view deep network instances training, respectively. In addition, four types of targets as shown in Table 5 with depression 30° are used to generate the testing

samples. Then, 2000 samples for each class are randomly selected for ATR performance evaluation in this experiment.

Table 5. Raw SAR images selection under EOC-D.

Training				Testing		
Types	**3-Views Instance**	**4-Views Instance**	**Depression**	**Types**	**3-/4-Views Instances**	**Depression**
2S1	100	75		2S1	288	
BRDM2	100	75	17°	BRDM2	287	30°
T72sn-132	78	58		T72-A64	288	
ZSU23/4	100	75		ZSU23/4	288	

The recognition results of the EOC-D test are shown in Table 6, and it can be seen that the three and four input view network instances can obtain good recognition results. The top recognition rate of the proposed network instance can reach more than 97.00%, and the recognition rates for all instances are higher than 95.00%. From an EOC-D test, we can see that the training data have a constant depression angle; however, the test results for the multiview deep network can still have relatively stable recognition performances under a large depression variation condition.

Table 6. Confusion Matrix of multiview deep convolutional neural networks under EOC-D.

Network Instances	Class	2S1	BRDM2	T72-A64	ZSU23/4	Recognition Rate (%)
3-Views	2S1	99.05	0.35	0.55	0.05	95.61
	BRDM2	0.00	100.00	0.00	0.00	
	T72-A64	13.40	1.30	84.95	0.35	
	ZSU23/4	0.60	0.00	0.95	98.45	
4-Views	2S1	96.40	0.10	3.50	0.00	97.12
	BRDM2	0.00	99.90	0.10	0.00	
	T72-A64	0.35	0.00	92.55	7.10	
	ZSU23/4	0.00	0.00	0.35	99.65	

Next, we will evaluate the performances of the proposed network under different target configurations and versions' testing conditions. The targets for training and testing have different components such as extra fuel tanks under the target configuration variation (EOC-C) test, while the version variation (EOC-V) test includes some structure difference among the training and testing targets, such as the rotation of the tank turret and so on. All of these conditions will add difficulties to accurate recognition but could be encountered in real applications.

In this experimental setup, there are four types of ground targets with depression 17° that are selected as raw SAR images, and their type and number for each instance are listed in Table 7. Thus, 14,445 and 11,380 multiview SAR image sequences are generated for the three and four input view instances training, respectively. The raw SAR images selection for testing sample generation under EOC-C and EOC-V are also listed in Table 8. Then, we randomly select 2000 samples for each target type variation from these generated testing data for performance evaluations.

The recognition results of EOC-C and EOC-V of three and four input view networks are shown in Tables 9 and 10. It is worth noting that the columns of the tables correspond to the four predicted classes, and the rows in these two confusion matrices denote the actual target type with configuration or version variations.

Table 7. Raw SAR images selection in the training phase under EOC-C and EOC-V.

Types	3-Views Instance	4-Views Instance	Depression
BMP2sn-9563	78	59	
BTR70	78	59	17°
T72sn-132	78	58	
BRDM2	100	75	

Table 8. Raw SAR images selection in the testing phase under EOC-C and EOC-V.

Operating Condition	Types	Number	Depression
EOC-C	BMP2sn-c21	429	17° and 15°
	BMP2sn-9566	428	
	T72sn-812	426	
	T72-A04	573	
	T72-A05	573	
	T72-A07	573	
	T72-A10	567	
EOC-V	T72sn-s7	419	17° and 15°
	T72-A32	572	
	T72-A62	573	
	T72-A63	573	
	T72-A64	573	

Table 9. Confusion Matrix of multiview deep convolutional neural networks under EOC-C.

Network Instances	Class	BMP2	BTR70	T72	BRDM2	Recognition Rate (%)
3-Views	BMP2sn-c21	98.55	0.05	1.25	0.15	96.49
	BMP2sn-9566	93.00	0.15	6.15	0.70	
	T72sn-812	3.45	0.05	95.85	0.65	
	T72-A04	7.65	0.15	91.50	0.70	
	T72-A05	0.25	0.05	99.50	0.20	
	T72-A07	1.50	0.00	97.90	0.60	
	T72-A10	0.50	0.10	99.15	0.25	
4-Views	BMP2sn-c21	99.05	0.00	0.85	0.10	97.84
	BMP2sn-9566	97.40	0.35	2.20	0.05	
	T72sn-812	2.55	0.05	96.90	0.50	
	T72-A04	3.45	0.10	95.90	0.55	
	T72-A05	1.90	0.00	97.90	0.20	
	T72-A07	1.65	0.00	98.25	0.10	
	T72-A10	0.35	0.05	99.50	0.10	

From Table 9, it can be observed that the recognition rates of the two network instances are higher than 96.00% and 97.00%, respectively. It shows that our proposed multiview network can achieve excellent ATR performance when the testing targets have different configurations.

Table 10 shows the recognition performances of the two network instances under EOC-V test. It can be seen that the proposed network with three input views can achieve a recognition rate over 96.00% in this experiment. In addition, with the input views of the network instance increasing to four, the recognition rate can rise to 99.00%. These experimental results above have proven that the proposed multiview deep feature learning network can obtain outstanding recognition performances under different ATR operating conditions.

Table 10. Confusion Matrix of multiview deep convolutional neural networks under EOC-V.

Network Instances	Class	BMP2	BTR70	T72	BRDM2	Recognition Rate (%)
3-Views	T72sn-s7	4.65	0.95	94.30	0.10	96.30
	T72-A32	0.70	0.00	99.20	0.10	
	T72-A62	0.60	0.05	98.60	0.75	
	T72-A63	3.95	0.00	95.15	0.90	
	T72-A64	4.85	0.15	94.25	0.75	
4-Views	T72sn-s7	0.20	0.00	99.75	0.05	99.10
	T72-A32	0.15	0.00	99.75	0.10	
	T72-A62	0.40	0.05	98.95	0.60	
	T72-A63	1.55	0.10	98.35	0.00	
	T72-A64	0.75	0.00	98.70	0.55	

4.5. ATR Performance Comparison

In this subsection, we compare the multiview deep feature learning network with six other methods which have been widely cited or recently published in SAR ATR. These

ATR methods for performance comparison are adaptive boosting (AdaBoost) [29], iterative graph thickening (IGT) [20], conditional Gaussian model (CGM) [30], joint sparse representation (JSR) [41], sparse representation-based classification (SRC) [42], and a multiview deep convolutional neural network (MDCNN) [45]. AdaBoost constructs an effective classifier as a linear combination of base classifiers for SAR ATR. IGT is a two-stage ATR framework applied in SAR images based on probabilistic graphical models. CGM is a good SAR ATR classification method based on conditional Gaussian models. In addition, JSR and SRC are two novel multiview SAR ATR methods based on a sparse representation theory, and MDCNN is a deep learning multiview SAR ATR method.

The recognition rates under SOC and EOC for each ATR method are listed in Table 11, and the results of ATR methods for comparison are cited from the related literature [20,30,41,42,45]. The proposed multiview deep feature learning networks with three and four input views are denoted as 3-VDFLN and 4-VDFLN, respectively. In addition, the ATR performance of the proposed deep network with just one view denoted as 1-VDFLN is also tested here as a classic counterpart. Table 11 shows that the accuracy rates of all the ATR methods are higher than 92.00% under SOC, but the performances of those SAR ATR methods are both different under SOC and EOC tests. It can be seen that, due to extracting much classification information from multiview SAR images, the recognition accuracy rates of the ATR methods with multiview inputs are generally higher than that with single-view approaches, especially under EOC tests. The comparison experiment results demonstrate that the proposed multiview deep feature learning networks have superior recognition performance in both SOC and EOC tests over the six other SAR ATR methods.

Table 11. Recognition performance of various SAR ATR methods.

ATR Methods	Recognition Rates (%)			
	SOC	EOC-D	EOC-C	EOC-V
AdaBoost	92.00	78.00	-	82.00
IGT	95.00	80.00	-	85.00
CGM	97.18	79.00	81.22	80.00
JSR	94.69	96.50	-	-
SRC	98.94	-	96.78	-
MDCNN	98.52	94.61	95.45	95.46
1-VDFLN	96.58	89.40	93.67	91.81
3-VDFLN	99.30	95.61	96.49	96.30
4-VDFLN	99.62	97.12	97.84	99.10

All of the above experiments have shown the outstanding recognition capabilities of the proposed multiview deep feature learning network, and have manifested the reasonability and validity of the multiview SAR ATR processing framework as well.

5. Conclusions

A reasonable and valid ATR framework and an effective ATR method are the two important issues incorporated into the multiview SAR ATR domain. In this paper, a new processing framework for a multiview SAR ATR pattern has been presented firstly. Based on this framework, a novel ATR method with a multiview deep feature learning network has been presented and applied to multiview SAR ATR as well. Two kinds of crucial classification features, i.e., the intra-view and inter-view features existing in the multiview SAR images, have been learned thoroughly by our multiview deep neural network.

Extensive experimental results have shown that the proposed multiview SAR ATR method can achieve excellent recognition performances. Its recognition rates with three and four views can reach 99.30% and 99.62% under SOC in a ten classes problem, respectively. In addition, it can achieve superior recognition performances compared to existing SAR ATR methods under various operating conditions such as depression angles, configurations, and version variations. These good recognition capabilities of the proposed neural network have also demonstrated the reasonability and validity of the given multiview

SAR ATR processing framework in this paper. The subsequent research mainly consists of new multiview ATR networks design and performance tests under more complex operating conditions. Additionally, we will study how to improve the multiview SAR ATR performance with small training samples.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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: London, 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. In *Encyclopedia of Aerospace Engineering*; Wiley: Hoboken, NJ, USA, 2010.
7. Tait, P. *Introduction to Radar Target Recognition*; The Institution of Engineering and Technology: London, UK, 2005; Volume 18.
8. Clemente, C.; Pallotta, L.; Gaglione, D.; De Maio, A.; Soraghan, J.J. Automatic Target Recognition of Military Vehicles With Krawtchouk Moments. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 493–500. [CrossRef]
9. 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]
10. Novak, L.M.; Owirka, G.J.; Brower, W.S.; Weaver, A.L. The automatic target-recognition system in SAIP. *Linc. Lab. J.* **1997**, *10*, 187–202.
11. 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*, 71598. [CrossRef]
12. Kreithen, D.E.; Halversen, S.D.; Owirka, G.J. Discriminating targets from clutter. *Linc. Lab. J.* **1993**, *6*, 25–52.
13. 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]
14. Gao, G. An Improved Scheme for Target Discrimination in High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 277–294. [CrossRef]
15. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 643–654. [CrossRef]
16. 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]
17. He, Y.; He, S.Y.; Zhang, Y.H.; Wen, G.J.; Yu, D.F.; Zhu, G.Q. A Forward Approach to Establish Parametric Scattering Center Models for Known Complex Radar Targets Applied to SAR ATR. *IEEE Trans. Antennas Propag.* **2014**, *62*, 6192–6205. [CrossRef]
18. 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]
19. Zhang, H.; Nasrabadi, N.M.; Huang, T.S.; Zhang, Y. Joint sparse representation based automatic target recognition in SAR images. In Proceedings of the SPIE 8051, Algorithms for Synthetic Aperture Radar Imagery XVIII, Orlando, FL, USA, 4 May 2011; p. 805112.
20. 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]
21. Cho, J.H.; Park, C.G. Multiple Feature Aggregation Using Convolutional Neural Networks for SAR Image-Based Automatic Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1882–1886. [CrossRef]

22. 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. [CrossRef]
23. He, Z.; Xiao, H.; Gao, C.; Tian, Z.; Chen, S. Fusion of Sparse Model Based on Randomly Erased Image for SAR Occluded Target Recognition. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7829–7844 [CrossRef]
24. Cao, C.; Cao, Z.; Cui, Z. LDGAN: A Synthetic Aperture Radar Image Generation Method for Automatic Target Recognition. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3495–3508. [CrossRef]
25. Sun, Y.; Du, L.; Wang, Y.; Wang, Y.; Hu, J. SAR automatic target recognition based on dictionary learning and joint dynamic sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1777–1781. [CrossRef]
26. Novak, L.M.; Benitz, G.R.; Owirka, G.J.; Bessette, L.A. ATR performance using enhanced resolution SAR. In *Algorithms for Synthetic Aperture Radar Imagery III*; International Society for Optics and Photonics: Bellingham, WA, USA, 1996; Volume 2757, pp. 332–337.
27. Mishra, A.; Mulgrew, B. Bistatic SAR ATR using PCA-based features. In *Automatic Target Recognition XVI*; International Society for Optics and Photonics: Bellingham, WA, USA, 2006; Volume 6234, p. 62340U.
28. Liu, X.; Huang, Y.; Pei, J.; Yang, J. Sample Discriminant Analysis for SAR ATR. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2120–2124.
29. 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]
30. 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]
31. Ikeuchi, K.; Wheeler, M.D.; Yamazaki, T.; Shakunaga, T. Model-based SAR ATR system. In *Aerospace/Defense Sensing and Controls*; International Society for Optics and Photonics: Bellingham, WA, USA, 1996; pp. 376–387.
32. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
33. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef]
34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
35. 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. [CrossRef]
36. 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: Bellingham, WA, USA, 2000; pp. 567–578.
37. Brown, M.Z. Analysis of multiple-view Bayesian classification for SAR ATR. In *Algorithms for Synthetic Aperture Radar Imagery X*; International Society for Optics and Photonics: Bellingham, WA, USA, 2003; Volume 5095, pp. 265–274.
38. 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 Navig.* **2015**, *9*, 457–466. [CrossRef]
39. Vespe, M.; Baker, C.J.; Griffiths, H.D. Aspect dependent drivers for multi-perspective target classification. In Proceedings of the 2006 IEEE Conference on Radar, Verona, NY, USA, 24–27 April 2006; p. 5.
40. Huan, R.H.; Pan, Y. Target recognition for multi-aspect SAR images with fusion strategies. *Prog. Electromagn. Res.* **2013**, *134*, 267–288. [CrossRef]
41. Zhang, H.; Nasrabadi, N.M.; Zhang, Y.; Huang, T.S. Multi-view automatic target recognition using joint sparse representation. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 2481–2497. [CrossRef]
42. Ding, B.; Wen, G. Exploiting multi-view SAR images for robust target recognition. *Remote Sens.* **2017**, *9*, 1150. [CrossRef]
43. 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]
44. Bai, X.; Xue, R.; Wang, L.; Zhou, F. Sequence SAR Image Classification Based on Bidirectional Convolution-Recurrent Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9223–9235. [CrossRef]
45. 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]
46. Novak, L.M.; Owirka, G.J.; Netishen, C.M. Performance of a high-resolution polarimetric SAR automatic target recognition system. *Linc. Lab. J.* **1993**, *6*, 11–24.
47. Mittermayer, J.; Moreira, A.; Loffeld, O. Spotlight SAR data processing using the frequency scaling algorithm. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2198–2214. [CrossRef]
48. Lin, Y.; Hong, W.; Tan, W.; Wang, Y.; Xiang, M. Airborne circular SAR imaging: Results at P-band. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 5594–5597.
49. Principe, J.C.; Xu, D.; Fisher, J.W., III. Pose estimation in SAR using an information theoretic criterion. In *Algorithms for Synthetic Aperture Radar Imagery V*; International Society for Optics and Photonics: Bellingham, WA, USA, 1998; Volume 3370, pp. 218–229.
50. Gonzalez, R.C.; Woods, R.E.; Eddins, S.L. *Digital Image Processing Using MATLAB*; Pearson Education India: Tamil Nadu, India, 2004.
51. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Deep learning for precipitation nowcasting: A benchmark and a new model. *arXiv* **2017**, arXiv:1706.03458

52. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [CrossRef]

53. Srivastava, N.; Hinton, G.E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

54. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 8 March 2021).

55. 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: Bellingham, WA, USA, 1998; pp. 566–573.

56. Maaten, L.v.d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.