

Technical Note

A Method of SAR Image Automatic Target Recognition Based on Convolution Auto-Encode and Support Vector Machine

Yang Deng ^{1,2,*} and Yunkai Deng ^{1,2}

¹ Space Microwave Remote Sensing System Department, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China

² School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100039, China

* Correspondence: dengyang18@mails.ucas.edu.cn

Abstract: In this paper, a method of Synthetic Aperture Radar (SAR) image Automatic Target Recognition (ATR) based on Convolution Auto-encode (CAE) and Support Vector Machine (SVM) is proposed. Using SVM replaces the traditional softmax as the classifier of the CAE model to classify the feature vectors extracted by the CAE model, which solves the problem that the softmax classifier is less effective in the nonlinear case. Since the SVM can only solve the binary classification problem, and in order to realize the classification of the class objectives, the SVM were designed to achieve the classification of the input samples. After unsupervised training for CAE, the coding layer is connected with SVM to form a classification network. CAE can extract the features of the data by an unsupervised method, and the nonlinear classification advantage of SVM can classify the features extracted by CAE and improve the accuracy of the object recognition. At the same time, the high-accuracy identification of key targets is required in some special cases. A new initialization method is proposed, which initializes the network parameters by pretraining the key targets and changes the weights of different targets in the loss function to obtain better feature extraction, so it can ensure good multitarget recognition ability while realizing the high recognition accuracy of the key targets.

Keywords: target recognition; deep learning; CAE



Citation: Deng, Y.; Deng, Y.A. Method of SAR Image Automatic Target Recognition Based on Convolution Auto-Encode and Support Vector Machine. *Remote Sens.* **2022**, *14*, 5559. <https://doi.org/10.3390/rs14215559>

Academic Editors: Xiangrong Zhang, Yansheng Li, Lichao Mou, Licheng Jiao and Xu Tang

Received: 27 September 2022

Accepted: 1 November 2022

Published: 4 November 2022

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



Copyright: © 2022 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

SAR images are widely used in the military field. Target recognition is the important support of military reconnaissance and attack. ATR can improve the efficiency of target recognition, and researchers in related fields have proposed a large number of ATR algorithms for SAR images. Among them, the relatively simple nearest-neighbor algorithm [1] and, with a moderate degree of complexity, are the Principal Component Analysis (PCA) [2], Two-Dimensional PCA (2DPCA) [3], and Sparse Representation Classification (SRC) [4]. Those with a high complexity are SVM [5] and deep learning [6–8]. A deep learning-based SAR image ATR algorithm is the best performance method [7]. In the traditional ATR method, the most important step is to find the most representative features of the target, such as using a discrete wavelet analysis to extract the target features [9], and then using SVM as a classifier for classification. The Principal Component Analysis (PCA) is used as the feature extractor, and then, the features are input into the ART2 neural network for classification [10,11], applying non-negative matrix factorization features to SAR target recognition [12].

In recent years, the target recognition based on deep learning has achieved great success, and the application of deep learning in SAR target recognition has been widely concerned. Generally speaking, because of the small number of military target samples and the complexity of a deep neural network, it is easy to be overfitted in model training, which affects the accuracy of recognition. Data expansion of the training sample set

can alleviate overfitting to some extent, but it cannot be avoided completely. Li X et al. proposed a method based on dividing the Convolutional Neural Network (CNN) into a CAE and Shallow Neural Network (SNN) to accelerate the training [13], while it also causes the loss of the recognition rate and also has certain limitations on the image angle. Housseini A E. et al. proposed deep learning algorithms based on the convolutional neural network architecture [14], extracting trained filters from the CAE and using them in the CNN has given a good result in terms of the time of calculation, but the accuracy is not better than either one. Chen S. et al. presented a new all-convolutional networks (A-ConvNets) [15]; the experimental results on the MSTAR benchmark data can achieve an average accuracy of 99%, but under Extended Operating Conditions (EOC), only achieve more than 80%. Wagner et al. proposed to use CNN to extract the feature vector [16] and then input the feature vector into SVM for classification. With the presented training methods, a correct classification rate of 99.5% in the confusion matrix best forced decision case was achieved. However, the experimental results under other conditions were more general. In the above articles, the experiments involving the MSTAR (Moving and Stationary Target Acquisition and Recognition) database [17] were only identified for ten different types of military targets, and the targets of the same type were not studied in depth, and the results were not extensive. However, in the related research of SAR image ATR, researchers usually set their problem under more ideal conditions. For example, in the existing literature on SAR image ATR technology, most people assume (or by default) that the center of the target to be identified is in the center of the sample image. There is literature assuming that the orientation of the target in SAR images is known or can be estimated [7]. The actual SAR image ATR system usually includes three steps: detection, identification, and recognition [18]. In the target detection stage, it is necessary to find the location of the military target from the large scene, and usually, the detection result will have a certain error [19]. In the target identification stage, it is necessary to exclude the detected interference target and estimate the target pose and others. Due to the errors in both target detection and identification, it is difficult to ensure that the target is at the center of the sample image intercepted after detection and ensure that the identified posture is consistent with the actual situation. In addition, most of the applications based on starborne SAR images are used in geocoded data products, and the original SAR images will undergo a certain geometric deformation. For small image blocks in large scenes, this deformation can be roughly approximated using rotation transformations. Therefore, in order to reduce the dependence on the detection and discrimination accuracy, the SAR ATR algorithm closer to a real situation should have a high robustness in the target position and the image rotation.

In this paper, a method of SAR image ATR based on CAE and SVM is proposed. CAE can extract the features of the data by using an unsupervised method, and the nonlinear classification advantage of SVM can classify the features extracted by CAE to get a better classification effect. To improve the utility of the training model in real SAR environments, the effects of image rotation and center offset on the classification accuracy of ATR models in SAR images were also investigated. We also verified that the model has high classification accuracy for 12 different rotation angles and has ideal classification results for different center offset datasets. At the same time, in order to satisfy the higher recognition accuracy, the key object recognition method of CAE based on special initialization and improved loss function is proposed. While maintaining a high overall identification accuracy, it can further improve the recognition accuracy of the key targets and achieve a high accuracy identification of the specified targets.

2. Automatic Target Recognition Method Based on CAE and SVM

2.1. CAE Model Training

CAE is a kind of self-encoder, which uses convolution operation to realize feature extraction and the reconstruction of data (Figure 1). Generally speaking, the self-encoder only contains a full connection layer, while the CAE coding part contains a convolution

layer and maximum pool layer, and the decoding part contains convolution layer and upper sampling layer (upsampling). Contrary to the effect of the pooling layer, upsampling is a process of increasing the data dimensions. There are two common methods of upper sampling: one is nearest-neighbor interpolation, and the other is linear interpolation. The results of the encoder convolution are upsampled in the model and then convolution-processed. Each maximum pool operation corresponds to an upper sampling operation, and the results after the upper sampling are convolutional, and the reconstructed image can be obtained.

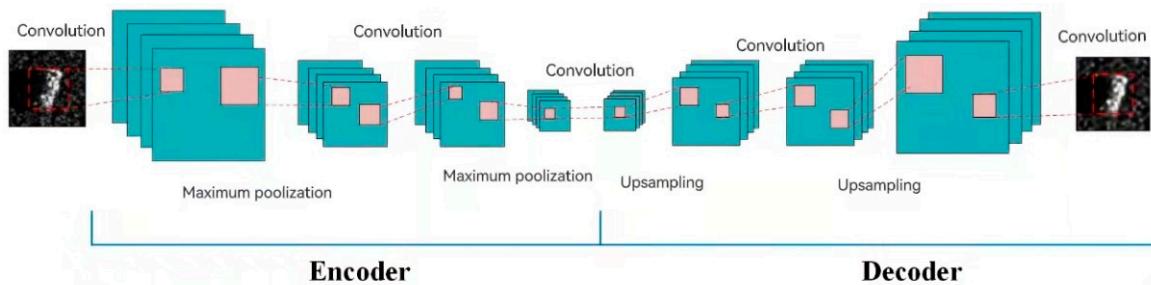


Figure 1. CAE model structure diagram.

Spatial locality is preserved by introducing convolution operations at each node. For a given input image matrix P , the encoding process can be expressed as:

$$e_n = \sigma(P \otimes F^n + b) \quad (1)$$

where σ is the activation function, \otimes is the two-dimensional convolution operation, F^n is the n convolution filter, and b is the encoder offset. In order to guarantee the spatial resolution, the input matrix P is zeroed. The decoding process can be expressed as:

$$\tilde{P}_n = \sigma\left(e \otimes \tilde{F}^n + \tilde{b}\right) \quad (2)$$

where \tilde{P}_n represents the reconstructed image, \tilde{F}^n represents the n deconvolution filter, and \tilde{b} represents the decoder offset.

The realization of the CAE model training is in accordance with the idea of a self-encoder. The difference between the reconstructed data and original data is used to train the CAE model, and the stable network parameters are obtained. Unsupervised training of the model with input sample data and updating of the model with a gradient descent to minimize the loss function can be defined as follows:

$$E = \sum_{i=1}^m \|P_i - \tilde{P}_i\|^2 \quad (3)$$

2.2. The Use of SVM

In recent years, ensemble learning based on multiple support vector machines (SVM) has been widely used in machine learning. Je et al. (Korea) integrated multiple SVM with the bagging method and successfully applied it to face recognition [20]. Frossyniotis et al. (UK) put forward an integrated SVM classification system based on supervised learning and unsupervised learning [21]. Chen et al. used a fuzzy fusion system to combine multiple SVM outputs [22], which made the algorithm more robust and stable. Since SVM has a strong nonlinear classification ability, many experimental research trials have been combined other classifiers or feature extraction algorithms with SVM, for example, the k-nearest neighbor algorithm and SVM fusion, Fuzzy Clustering Algorithm and SVM fusion [23], decision tree and SVM fusion [24], and CNN and SVM fusion [16].

In general, the output layer of the self-encoder model in the case of multiclassification uses the softmax function as the classifier. However, the effect of the softmax classifier is not ideal in the nonlinear case. In order to solve this problem, SVM is used instead of softmax as the classifier of the CAE model, and the feature vectors extracted from the CAE model are classified. The structure of CAE + SVM is shown in Figure 2.

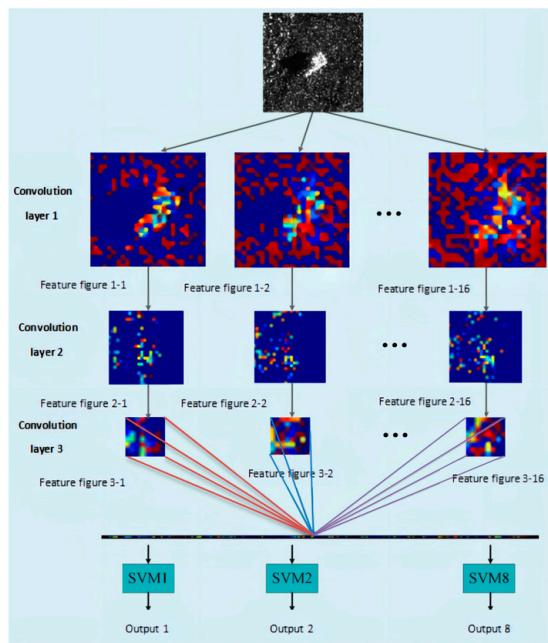


Figure 2. CAE + SVM model structure.

2.3. Target Identification Method

Military target recognition can be divided into two steps: detecting military targets from complex background and classifying military targets separated from the background.

SAR images usually contain complex background information, such as forests, urban areas, and water areas. The first step of ATR is to extract the military target from the complex background, which can be considered as a binary classification problem. The classification can be divided into three steps: training sample set acquisition, CAE model feature extraction, and SVM classification. Small images with military targets and backgrounds were selected as training samples. After that, unsupervised pretraining of the CAE model was carried out. In this process, the features of the training samples are extracted by the CAE model, and then, the extracted features are connected with SVM for supervised training. After training, the test samples are input into CAE for feature extraction, and the extracted features are input into SVM for classification. The samples are divided into two categories: samples containing military objectives and background samples. Target detection is accomplished by the classification of military objectives and complex backgrounds.

After the targets are extracted from the background, the extracted military targets are classified, and the classification is still realized by combining CAE and SVM. Different military targets are labeled, and the labeled military targets are grouped into the training sample set. The specific implementation method of the classification is the same as that described above.

2.4. Automatic Target Recognition Based on Improved Loss Function

By changing the initialization mode of the CAE network and improving the loss function, the military target recognition is realized with high recognition precision and good multitarget recognition ability. In order to achieve the above-mentioned goal, firstly, the model is pretrained by using the training samples of the key targets. The pretraining process is similar to the unsupervised training process by reducing the reconstruction error

between the output value and the input value. The network parameters after pretraining are the initialization parameters of the subsequent model training, so the process can also be considered a network initialization.

In order to learn better features from the training samples, the supervised training method is used to fine tune the parameters of the model. The softmax classifier is connected with the encoder part of the model, and then, the parameters of the model are optimized with labeled training samples. In general, the loss function can be defined as:

$$J(\theta) = \operatorname{argmin}_{\theta} \sum_{i=1}^N \|l_{r(i)} - \chi(f_{\theta}(x_i))\|^2 \quad (4)$$

where N is the number of training samples, $r(i)$ are the i samples that belong to category r , $l_{r(i)}$ is the corresponding category label, $\chi(\cdot)$ is the classification process, f_{θ} is the encoding process, and x_i is the input sample.

$$J'(\theta) = \operatorname{argmin}_{\theta} \sum_{i=1}^N \omega_{r(i)} \|l_{r(i)} - \chi(f_{\theta}(x_i))\|^2 \quad (5)$$

where ω_r represents the weight of the class r sample and satisfies $\sum_r \omega_r = 1$. The weight of the key target in the loss function can be increased by increasing the ω_r corresponding to the key target, and then, the ability of identifying the key target can be improved.

3. Experimental Results and Analysis

In this paper, the MSTAR database is used for experimental verification. The database is provided by the MSTAR project jointly funded by the US Defense Prereseach Program Agency (DARPA) and the US Air Force Laboratory (AFRL). The database contains SAR images of different military targets obtained by a 0.3-m resolution X-band cluster mode airborne SAR system at different pitch angles and 360 circular flight conditions. The database includes three datasets: T72_BMP2_BTR70, T72 variant, and mixed target. The mixed target dataset includes seven different categories of military targets, namely armored personnel carriers: BTR-60 and BRDM-2, rocket launcher: 2S1, bulldozer: D7, tank: T62, truck: ZIL-131, and air defense unit: ZSU-234, whose optical photos and SAR images are shown in Figure 3. The T72_BMP2_BTR70 dataset contains three types of targets with similar structures. Among them, BMP2 armored vehicles have three models: SN-C21, SN-9563, and SN-9566, and the T72 main battle tank has SN-132, SN-812, and SN-S7. The T72 variant dataset includes eight different models of the T72 main battle tank, namely, the A04, A05, A07, A10, A32, A62, A63, and A64.

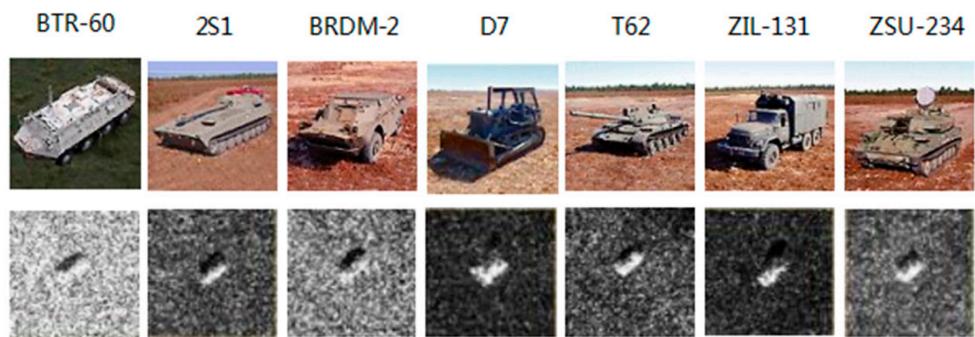


Figure 3. Military targets of the mixed target dataset.

The features of the target in the SAR images are very sensitive to the incident direction during shooting. Therefore, although the MSTAR database provides the SAR images taken at the 360° azimuth angle, it is impossible to describe the phenomenon of image rotation after shooting. The image sizes of the different targets in the original data were inconsistent,

but the targets were all located in the center of the image. The SAR image looks blurred and out of focus due to a lack of resolution and the noise created by the background and SAR data processing. Using the T72_BMP2_BTR70 dataset as an example, samples with a rotation transformation and a center offset were generated using the original images. Using the original image center as the reference point, the intercepted image block of size 80×80 was taken as the sample used in the experiment. To expand the dataset, the resulting training samples were manipulated by translation and rotation, as shown in Figure 4. The training sample set was expanded by 12 times for all the images of 30° in the range from 0° to 330° . The translation operation of the image pixels was also performed, shifting in the range of 9 by 9 pixels in the two-dimensional direction of the sample image with three pixels as the steps. This translation mode expands the training sample set 36 times. For feature extraction training using the previously introduced method, the training sample set is input into the CAE model for pretraining, encoding the input samples using the encoder, and used to reconstruct the input samples using the decoder. The above data amplification method was used in the pretraining, and 3000 samples of each class were randomly selected from the amplified dataset as the training samples. The parameters of the CAE model used in the experiment are set as shown in Figure 5. The encoder contains the four convolutional layers and the three maximum pooling layers, and the first three convolution layers are all connected behind them. The kernel function size of the four convolution layers is 6×6 , 5×5 , 5×5 , and 3×3 , and the number of kernel functions is successively 128, 64, 32, and 16. The decoder contains four convolutional layers, three upsampling layers, and the latter three convolution layers are all connected to an upsampling layer. The kernel function sizes of the four convolution layers are 3×3 , 5×5 , 5×5 , and 6×6 , and the number of kernel functions is successively 16, 32, 64, and 128. After CAE training, the reconstruction result of the target is as shown in Figure 6. The first line is the original target image, and the next line is the reconstructed image. The reconstruction loss is 0.12. It can be seen from the diagram that the CAE model after training can reconstruct the original image well and has good capability of feature extraction.

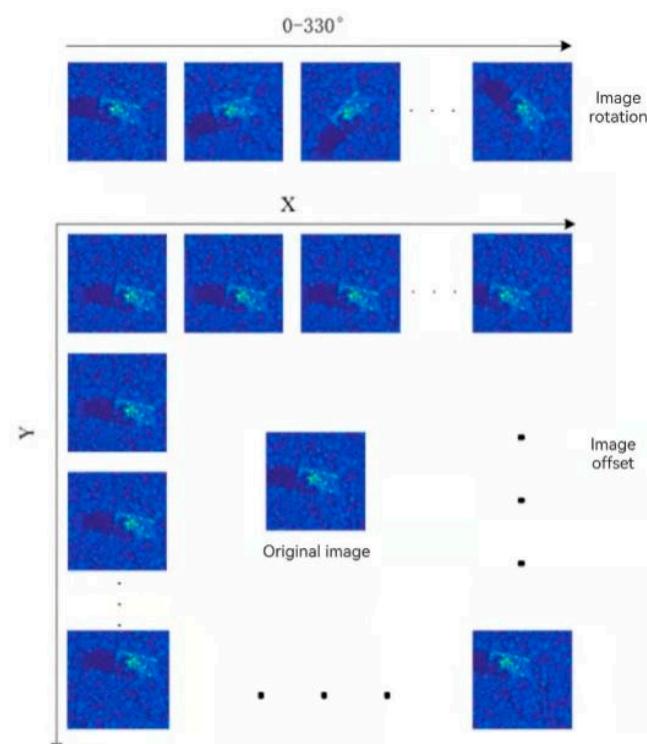


Figure 4. Dataset augmentation operation.

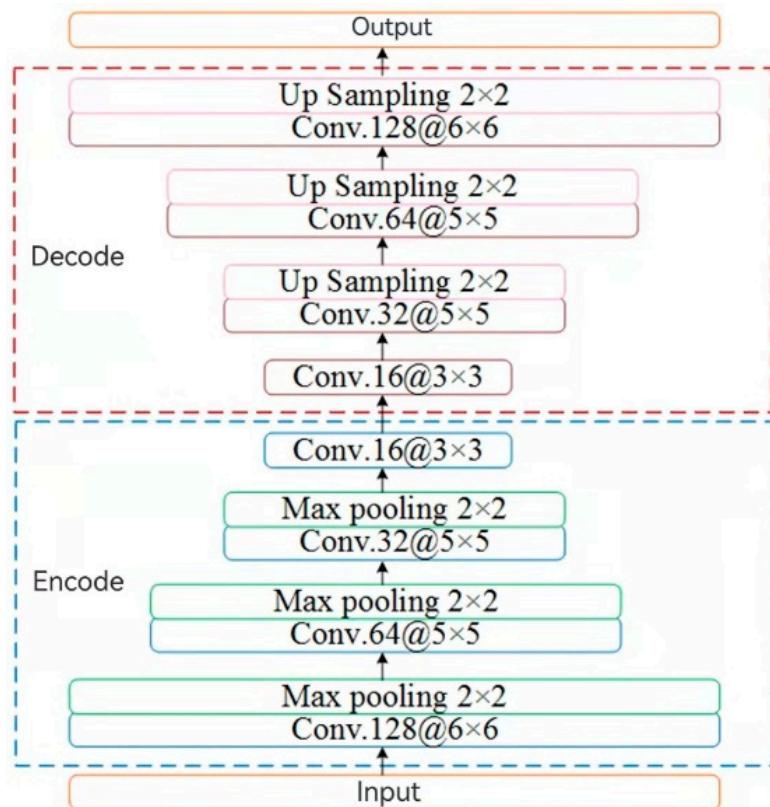


Figure 5. Model parameters setting.

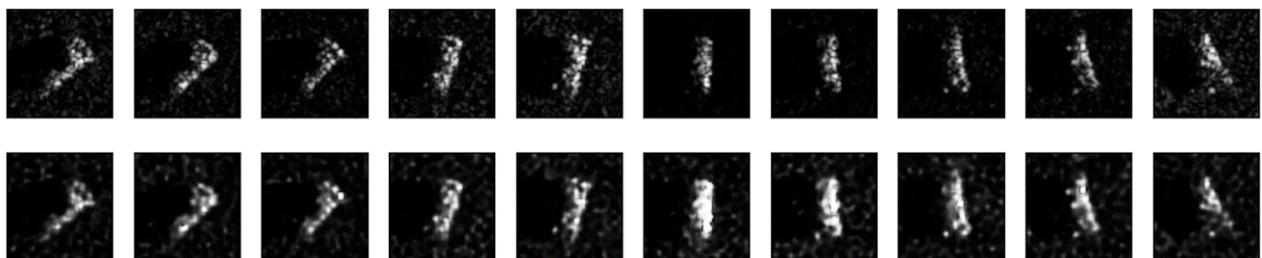


Figure 6. Original target images (up) and reconstructed target images with CAE (down).

After pretraining, the coding layer of the CAE model is connected with SVM to form a classification network; thus, the classification of military targets is realized. The identification results for the three types of targets in the T72_BMP2_BTR70 dataset are shown in Table 1. The recognition results of a class of targets in the table represent the statistics of the corresponding class after classification. As can be seen from the table, the classification accuracy of the three types of targets reached more than 98%. The target recognition accuracy of the original CAE method with softmax as the classifier are also given. The difference between the two methods is that the original CAE method uses softmax as the classifier for classification recognition after pretraining while the CAE and SVM fusion method uses SVM as the classifier. The average recognition rate of the original CAE method is 97.0% and that of the fusion method of CAE and SVM is 99.0%, which is 2% higher than that of the original CAE method. In addition, the classification accuracy of other target recognition methods such as SVM, Multilayer Perceptron (MLP) and AdaBoost are given in Table 2. It can be seen that the target recognition accuracy of the CAE and SVM fusion method proposed in this paper is better than that of the other methods given in the table. In order to verify the above conclusions, eight kinds of military targets in the T72 variant dataset, and seven kinds of military targets in the mixed target dataset were

classified and identified, respectively, in the experiment, and the recognition accuracy of each military target can reach above 90%.

Table 1. Target identification results (T72_BMP2_BTR70 dataset).

Test Target	Recognition Result			Accuracy Rate
	BTR70	BMP2	T72	
BTR70	198	1	0	99.5%
BMP2	2	197	1	98.5%
T72	1	1	198	99.0%
Average accuracy			99.0%	

Table 2. Target recognition results with different methods (T72_BMP2_BTR70 dataset).

Method	Recognition Accuracy
SVM	88.5%
MLP	87.0%
Adaboost	90.83%
CNN	95.50%
CAE	97.0%
CAE + SVM	99.0%

The model was trained using a training set with both a rotation angle and a center offset of 0. The trained models were used to predict the training and test sets with different rotation angles and different central offsets. The test on the training set can verify that, in the same original data, it is only affected by the accuracy of the model. The classification results for samples with different rotation angles are shown in Table 3.

Table 3. Classification accuracy corresponding to the different rotation angles (T72_BMP2_BTR70 dataset).

Rotation Angle (degree)	30	60	90	120	150	180
training set	99.3%	97.9%	99.2%	98.7%	99.9%	99.2%
test set	96.4%	94.2%	96.7%	97.1%	98.0%	96.3%
Rotation Angle (degree)	210	240	270	300	330	360
training set	99.1%	1	98.7%	97.9%	99.3%	1
test set	95.9%	98.0%	95.4%	94.5%	96.7%	98.1%
Average accuracy				96.4%		

It can be seen that the model obtained by the synthetic training set has a high classification accuracy for all 15 different rotation angles. Although the classification accuracy of the model in the training set is higher than that of the test set, the 96.4% test set accuracy is enough to show that the model is effective and stable in ideal conditions (without rotation and translation).

The test results of this model on sample sets with different offsets are shown in Table 4. The classification accuracy of the model for the training and test sets with a central offset is correlated with the offset. Overall, the larger the central offset, the lower the prediction accuracy of the model is, and the accuracy is slightly decreased as the offset increases.

Table 4. Classification accuracy at different offsets (T72_BMP2_BTR70 dataset).

X/Y (Pixel)	-9	-6	-3	0	3	6	9
-9	95.6%	94.7%	94.4%	96.9%	97.5%	95.7%	95.0%
-6	95.7%	95.5%	94.2%	95.9%	96.7%	95.4%	94.7%
-3	94.5%	93.9%	92.0%	94.1%	93.6%	93.7%	93.9%
0	94.7%	94.9%	95.8%	97.8%	95.5%	96.5%	94.8%
3	94.5%	94.0%	95.6%	97.5%	96.8%	96.6%	95.0%
6	94.4%	94.1%	96.5%	96.4%	95.8%	95.6%	95.6%
9	92.3%	93.3%	96.0%	96.3%	95.4%	94.9%	94.5%

Since high-resolution SAR images have smaller pixel blocks required to contain a single military target, some geometric transformations of large scene images (such as oblique conversion) have less impact on the military target image block. Even if the image of a large scene is geocoded, the deformation of the image block of the military target can be roughly approximate by the rotation transformation. On the other hand, the offset of the center as a major error in target extraction also has implications for target classification. Therefore, the image rotation and center offset can be taken as the important factors affecting the performance of starborne SAR, system army, and matter and event target ATR system. The robustness of recognition algorithms in rotation and translation can improve the performance of ATR systems in real application environments. Through the above experiments, we can see that the proposed model has a high classification accuracy for all 12 different rotation angles, and the classification results for the datasets with different central offsets are also very ideal.

In the previous experiments, only the slice image containing the target is processed, and only the target classification process in ATR is involved. The actual ATR process also includes the step of target detection, because the SAR scene image generally contains complex objects such as trees, buildings, grasslands, water bodies, etc., so the goal needs to be separated from the complex background first. Then, the target slice images are classified to recognize different target categories. In this paper, two-step improved CAE algorithm is used to realize the object end-to-end recognition. Firstly, the improved CAE algorithm is used to realize the binary classification of the scene image; that is, the scene is divided into target and background, and the target is extracted from the scene. Then, multiclassification recognition is carried out on the extracted target image. The open MSTAR dataset provides a 1748×1478 pixel image of the scene and the target slice. However, these scene images do not contain military targets. For experimental verification, a 128×128 pixel object slice image is embedded into the background image, as shown in Figure 7a, which contains the background and three types of military objects in the T72_BMP2_BTR70 dataset. This method is reasonable, because both the background image and the target slice image are obtained by the same SAR system with a resolution of 0.3 m.

The target detection can be realized by binary classification, in which the fusion method of CAE and SVM is adopted, the target and background samples are used as training samples in the process of target detection, and the features are extracted through the trained CAE model; then, SVM is used as a classifier for classification. The background sample is selected randomly from the background image, and the target sample is selected from the same dataset as before. An example of a training sample set is shown in Figure 8. The target contains different types of military targets in various poses. The background sample contains a variety of ground objects, such as trees, meadows, buildings, and farmland.

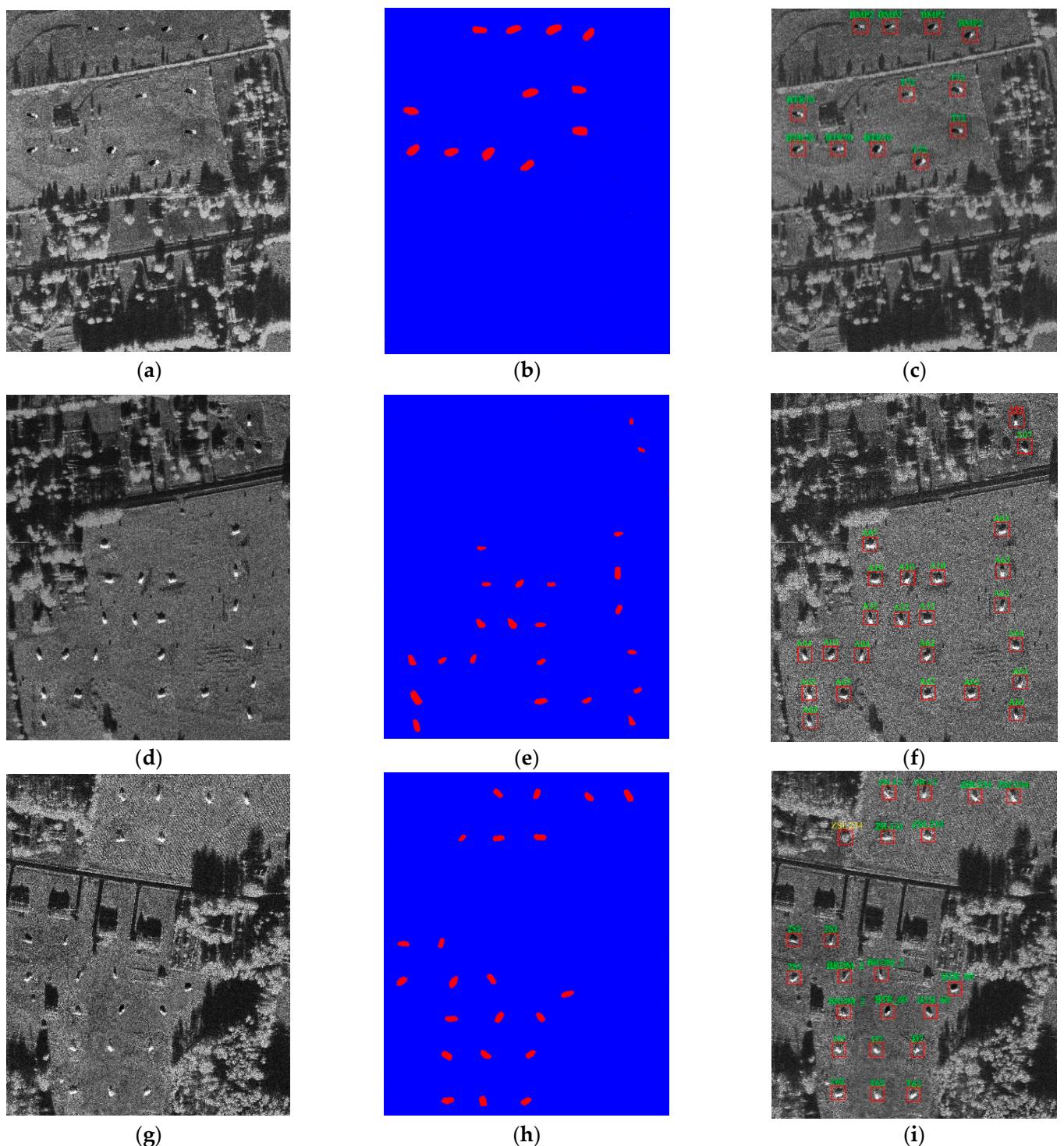


Figure 7. Target detection and recognition results. (a) Scene Image 1 containing the target. (b) Target detection results. (c) Target identification results. (d) Scene Image 2 containing the target. (e) Target detection results. (f) Target identification results. (g) Scene Image 3 containing the target. (h) Target detection results. (i) Target identification results.

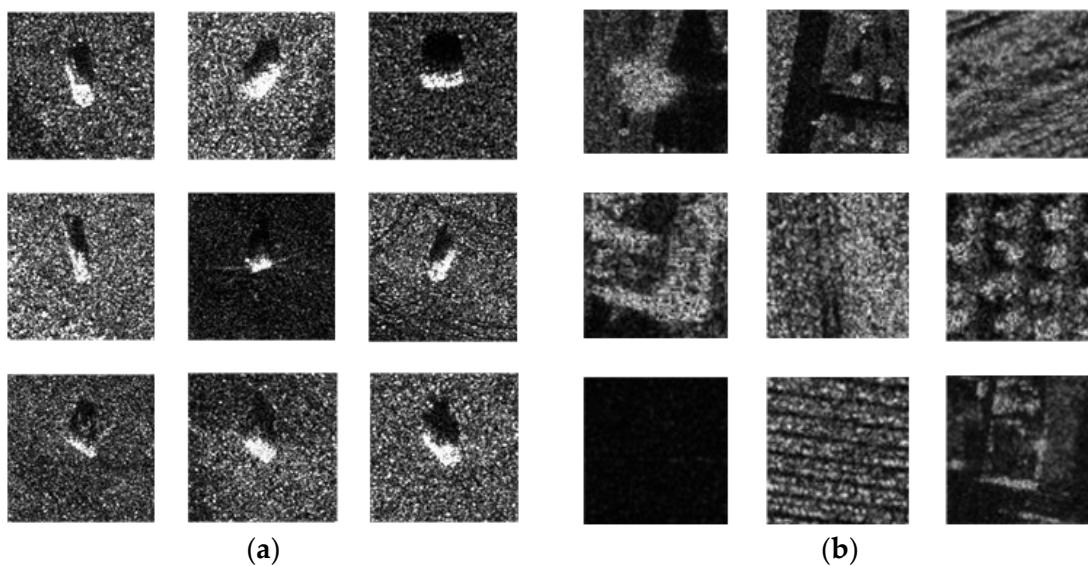


Figure 8. Training samples. (a) Target sample. (b) Background sample.

The trained binary CAE model can be used as a target detector to recognize military targets in complex scenes by a sliding window. The output of the classification is the probability of belonging to two categories. Here, 0.8 is chosen as the threshold value, and the area where the probability of belonging to a military target is greater than 0.8 is judged to be a military target; the three types of target detection results from the T72_BMP2_BTR70 dataset are shown in Figure 7b. Red is the location area where the target may exist, and blue is the background area. According to the detection results, the slice image corresponding to the target area can be extracted from the scene, in which the 80×80 target image is extracted with the target as the slice center. Then, the separated object samples can be input into the previously trained multiclass CAE model to get the object category. If the T72, BMP2, and BTR70 objects detected in Figure 7a are input into the corresponding CAE model that has been trained before, the target categories can be obtained. The final result is shown in Figure 7c. The red box is the location of the target in the diagram, and the green text is the category of the target.

In the same way, the experiments are carried out on the scene in Figure 7d of eight kinds of targets with the T72 variant and the scene in Figure 7g of seven kinds of targets with mixed targets. The results of target detection are shown in Figure 7e,h, respectively. The final identification result of the eight target classes of the T72 variant is shown in Figure 7f, where the target class in the upper right corner is marked in red, indicating that the target is misclassified. The final recognition result of the seven mixed targets is shown in Figure 7i, where the target category in the upper left corner is marked with yellow, indicating that the area is the background and is misclassified as the target. Although the result of recognition is misclassified and a false alarm, the result of statistical experiment on a large number of images shows that the accuracy of the whole recognition is still very high, which proves the validity of the fusion method of CAE and SVM.

The precise identification of the special object by changing the CAE network initialization and improving the loss function is verified by the MSTAR public database. This paper takes eight kinds of military targets of the T72 variant as an example to carry out target recognition. In order to study the influence of the improved loss function method on different targets, two targets, A04 and A10, are set as the key targets, and the other targets are the common targets.

In the experiment, the results of target recognition without the above initialization and the improvement of the loss function are studied, as shown in Table 5. The identification accuracy of the key targets is not better than the other categories of targets, and failed to achieve the key identification of specific targets. In order to improve the identification

accuracy of the key targets, the model was first trained by using the key target training samples to pretrain to achieve the model initialization, but the improved loss function was not adopted. The resulting target identification results are shown in Table 6. In the experiment, the effect of the improved loss function on the accuracy of key target recognition is further studied. After model initialization and supervised training, the improved loss function is adopted. The result of target recognition is shown in Table 7.

Table 5. Target recognition results (T72 dataset).

Test Target	Recognition Result								Accuracy Rate
	A04	A05	A07	A10	A32	A62	A63	A64	
A04	184	4	3	3	1	2	2	1	92.0%
A05	3	190	1	2	1	0	2	0	95.0%
A07	2	3	189	1	2	2	0	1	94.5%
A10	3	1	3	182	5	2	3	1	91.0%
A32	2	0	0	1	195	1	1	0	97.5%
A62	4	0	5	2	2	184	1	2	92.0%
A63	1	2	2	3	2	3	187	0	93.5%
A64	2	1	1	2	2	1	2	189	94.5%
Average accuracy									93.75%

Table 6. Target recognition results with pretraining (T72 dataset).

Test Target	Recognition Result								Accuracy Rate
	A04	A05	A07	A10	A32	A62	A63	A64	
A04	186	3	2	3	1	2	2	1	93.0%
A05	3	190	1	2	1	0	2	1	95.0%
A07	2	3	188	1	2	2	1	1	94.0%
A10	2	1	3	185	4	2	2	1	92.5%
A32	2	1	0	1	194	1	1	0	97.0%
A62	4	0	5	2	2	184	1	2	92.0%
A63	1	2	3	3	2	3	185	1	92.5%
A64	2	1	1	3	2	1	2	188	94.0%
Average accuracy									93.75%

Table 7. Target recognition results with pretraining and loss function improvement.

Test Target	Recognition Result								Accuracy Rate
	A04	A05	A07	A10	A32	A62	A63	A64	
A04	192	1	2	0	1	2	1	1	96.0%
A05	3	186	2	2	3	1	2	1	93.0%
A07	3	3	185	2	3	2	1	1	92.5%
A10	1	1	2	191	2	0	1	2	95.5%
A32	2	3	2	2	187	2	1	1	93.5%
A62	4	2	4	2	3	181	2	2	90.5%
A63	2	3	3	3	1	3	183	2	91.5%
A64	2	3	2	3	2	3	3	182	91.0%
Average accuracy									92.94%

The experimental results show that the recognition accuracy of the key target is improved greatly after the model initialization and the improvement of the loss function.

Although the recognition accuracy of ordinary targets decreases slightly, it is still above 90%, and the overall recognition accuracy of the target is kept at a good level.

To further verify the above conclusions, the seven categories of mixed targets were experimentally verified, and the 2s1 and T62 targets were selected as the key targets. The first set of experiments investigated without network initialization and improved loss function, and the target identification results are shown in Table 8, it can be seen that the key identification of specific goals has not been achieved. The second set of experiments used the key target samples to achieve the model initialization and did not adopt the improved loss function, and the target identification results are shown in Table 9, the recognition accuracy of the key targets 2s1 and T62 was improved by 1% compared to the previous recognition results, while the ordinary targets were slightly decreased, with the average accuracy decreasing by 0.29%. The third set of experiments further investigated the effect of the improved loss function on the recognition accuracy of the two key targets of 2s1 and T62. The target identification results are shown in Table 10, and the recognition accuracy of the key targets 2s1 and T62 was improved 3% and 4% compared to the original method, the average recognition accuracy decreased by about 1.22% relative to the results of the original method. The above experimental results show that to initialize the model and improve the loss function can further improve the recognition accuracy of key targets while maintaining the high overall recognition accuracy and play a great role in the key identification and extraction of special targets.

Table 8. Target recognition results (mixed targets dataset).

Table 9. Target recognition results with pretraining (mixed targets dataset).

Table 10. Target recognition results with pretraining and loss function improvement (mixed targets dataset).

Test Target	Recognition Result							Accuracy Rate
2s1	2s1 195	BRDM_2 1	BTR_60 1	D7 0	T62 2	ZIL131 0	ZSU234 1	97.5%
BRDM_2	2	183	6	1	2	2	4	91.5%
BTR_60	3	5	182	2	3	2	3	91.0%
D7	1	2	2	189	0	4	2	94.5%
T62	3	1	1	0	194	0	1	97.0%
ZIL131	1	3	2	3	1	187	3	93.5%
ZSU234	3	4	5	3	2	3	180	90.0%
Average accuracy				93.57%				

4. Conclusions

In this paper, an unsupervised training method for CAE was proposed, in which the coding layer and SVM connected to form a classification network. SVM with good nonlinear classification performance can improve the accuracy of target recognition. Experimental results showed that, compared with the softmax classifier, using SVM as the CAE classifier can improve the accuracy of target recognition. It also verified that the model has high classification accuracy for 12 different rotation angles and has ideal classification results for different center offset datasets. At the same time, a key target recognition method based on CAE was proposed, which can improve the recognition accuracy of key targets by initializing the model with key target training samples, and by changing the weight of different targets in the loss function, it can further improve the recognition accuracy of the key target. The experimental results showed that the ATR method based on CAE and SVM is a powerful tool for target recognition and detection using SAR images.

Author Contributions: Conceptualization, Y.D. (Yang Deng) and Y.D. (Yunkai Deng); methodology, Y.D. (Yang Deng); software, Y.D. (Yang Deng); validation, Y.D. (Yang Deng); formal analysis, Y.D. (Yang Deng); investigation, Y.D. (Yang Deng); resources, Y.D. (Yunkai Deng); data curation, Y.D. (Yang Deng); writing—original draft preparation, Y.D. (Yang Deng); writing—review and editing, Y.D. (Yang Deng); visualization, Y.D. (Yang Deng); supervision, Y.D. (Yunkai Deng); project administration, Y.D. (Yunkai Deng); funding acquisition, Y.D. (Yunkai Deng). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

- Rosenbach, K.; Schiller, J. Identification of aircraft on the basis of 2-D radar images. In Proceedings of the International Radar Conference, Alexandria, VA, USA, 8–11 May 1995; pp. 405–409.
- Zhang, R.; Hong, J.; Ming, F. An improved PCA based features for SAR ATR. In Proceedings of the IET International Radar Conference, Guilin, China, 20–22 April 2009; pp. 1–3.
- Pei, J.; Huang, Y.; Liu, X.; Yang, J.; Cao, Z.; Wang, B. 2DPCA-based Two-dimensional Maximum Interclass Distance Embedding for SAR ATR. In Proceedings of the 2013 International Conference on Communications, Circuits and Systems (ICCCAS), Chengdu, China, 15–17 November 2013; Volume 1, pp. 267–270.
- Zhang, H.; Nasrabadi, N.; Zhang, Y.; Huang, T.S. Multi-View Automatic Target Recognition using Joint Sparse Representation. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 2481–2497. [[CrossRef](#)]
- Zhao, Q.; Principe, J.C.; Brennan, V.; Xu, D.; Wang, Z. Synthetic aperture radar automatic target recognition with three strategies of learning and representation. *Opt. Eng.* **2000**, *39*, 1230–1244. [[CrossRef](#)]
- Wagner, S. Combination of convolutional feature extraction and support vector machines for radar ATR. In Proceedings of the 17th International Conference on Information Fusion (FUSION), Salamanca, Spain, 7–10 July 2014; pp. 1–6.

7. Wagner, S. Morphological Component Analysis in SAR images to improve the generalization of ATR systems. In Proceedings of the 2015 3rd International Workshop on Compressed Sensing Theory and Its Applications to Radar, Sonar and Remote Sensing (CoSeRa), Pisa, Italy, 17–19 June 2015; pp. 46–50.
8. Chen, S.; Wang, H. SAR target recognition based on deep learning. In Proceedings of the 2014 International Conference on Data Science and Advanced Analytics (DSAA), Shanghai, China, 30 October–1 November 2014; pp. 541–547.
9. Cheng, G.; Zhao, W.; Pan, J. Research on MSTAR SAR target recognition based on wavelet analysis and support vector machine. *J. Image Graph.* **2009**, *14*, 317–322.
10. Ye, X.; Gao, W.; Wang, Y. Research on SAR images recognition based on ART2 neural network. In Proceedings of the 7th IEEE Conference on Industrial Electronics and Applications (ICIEA), Singapore, 18–20 July 2012; pp. 1888–1891.
11. Ni, J.C.; Lei, X.Y. SAR automatic target recognition based on a visual cortical system. *Int. Congr. Image Signal Process.* **2014**, *2*, 778–782.
12. Huan, R.H.; Pan, Y.; Mao, K.J. SAR image target recognition based on NMF feature extraction and Bayesian decision fusion. In Proceedings of the 2010 Second IITA International Conference on Geoscience and Remote Sensing (IITA-GRS), Qingdao, China, 28–31 August 2010; pp. 496–499.
13. Li, X.; Li, C.; Wang, P.; Men, Z.; Xu, H. SAR ATR based on dividing CNN into CAE and SNN. In Proceedings of the 2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Singapore, 1–4 September 2015.
14. Housseini, A.E.; Toumi, A.; Khenchaf, A. Deep Learning for Target recognition from SAR images. In Proceedings of the 2017 Seminar on Detection Systems Architectures and Technologies (DAT), Algiers, Algeria, 20–22 February 2017.
15. 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](#)]
16. Wagner, S.A. SAR ATR by a combination of convolutional neural network and support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2016**, *52*, 2861–2872. [[CrossRef](#)]
17. Keydel, E.R.; Lee, S.W.; Moore, J.T. MSTAR extended operating conditions: A tutorial. *Algorithms Synth. Aperture Radar Imag. III* **1996**, *2757*, 228–242.
18. Jianxiong, Z.; Zhiguang, S.; Xiao, C.; Qiang, F. Automatic target recognition of SAR images based on global scattering center model. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3713–3729. [[CrossRef](#)]
19. Park, J.-I.; Park, S.-H.; Kim, K.-T. New Discrimination Features for SAR Automatic Target Recognition. *IEEE Trans. Geosci. Remote Sens.* **2013**, *10*, 476–480. [[CrossRef](#)]
20. Je, H.M.; Kim, D.; Bang, S.Y. Human Face Detection in Digital Video Using SVM Ensemble. *Neural Process. Lett.* **2003**, *17*, 239–252. [[CrossRef](#)]
21. Frossyntiotis, D.S.; Stafylopatis, A. A Multi-SVM Classification System. In Proceedings of the Second International Workshop on Multiple Classifier Systems, Cambridge, UK, 2–4 July 2001; pp. 198–207.
22. Chen, X.; Harrison, R.; Zhang, Y.Q. Multi-SVM Fuzzy Classification and Fusion Method and Applications in Bioinformatics. *J. Comput. Theor. Nanosci.* **2005**, *2*, 534–542. [[CrossRef](#)]
23. Zhang, H.; Shi, W.; Liu, K. Fuzzy-Topology-Integrated Support Vector Machine for Remotely Sensed Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 850–862. [[CrossRef](#)]
24. Moustakidis, S.; Mallinis, G.; Koutsias, N.; Theocharis, J.B. SVM Based Fuzzy Decision Trees for Classification of High Spatial Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 149–169. [[CrossRef](#)]