PAPER

An Algorithm Based on Distance Measurement for SAR Image Recognition

Yuxiu LIU[†], **Na WEI**[†], ***and*** **Yongjie LI**[†a)], ***Nonmembers***

SUMMARY In recent years, deep convolutional neural networks (CNN) have been widely used in synthetic aperture radar (SAR) image recognition. However, due to the difficulty in obtaining SAR image samples, training data is relatively few and overfitting is easy to occur when using traditional CNNS used in optical image recognition. In this paper, a CNN-based SAR image recognition algorithm is proposed, which can effectively reduce network parameters, avoid model overfitting and improve recognition accuracy. The algorithm first constructs a convolutional network feature extractor with a small size convolution kernel, then constructs a classifier based on the convolution layer, and designs a loss function based on distance measurement. The networks are trained in two stages: in the first stage, the distance measurement loss function is used to train the feature extraction network; in the second stage, cross-entropy is used to train the whole model. The public benchmark dataset MSTAR is used for experiments. Comparison experiments prove that the proposed method has higher accuracy than the state-of-the-art algorithms and the classical image recognition algorithms. The ablation experiment results prove the effectiveness of each part of the proposed algorithm.

key words: *SAR image recognition, convolutional neural networks, distance measurement, all convolutional network*

1. Introduction

Synthetic aperture radar (SAR) can realize high-resolution microwave remote sensing imaging by using the principle of synthetic aperture, which has advantages of all-sky, all-weather and strong penetration etc., and has important application value in Marine monitoring, environmental analysis, military reconnaissance and geological survey [1], [2]. However, the principle of SAR imaging determines that SAR images have strong speckle noise and geometric distortion [3], [4], which poses challenges to SAR image interpretation. Automatic Target Recognition (ATR) which integrates image detection, sign extraction, and image recognition processes, is one of the key technologies for achieving automatic interpretation of SAR images [5]–[7].

Image recognition, as the last crucial aspect of ATR, has been the focus of SAR imaging research. Traditional SAR image recognition methods mainly include two steps: feature extraction and classification recognition. The commonly used extracted features include geometric features, projection features and scattering features [8], [9], and the general classification and recognition algorithms include K-Nearest Neighbor classifier (KNN) [10], support vector machine (SVM) [11], sparse representation classifier [11], [12], etc. Traditional SAR image recognition methods have achieved good results and brought certain research progress to SAR ATR. However, the effectiveness of these methods requires experts to manually extract features, which is complicated, inefficient and poor robustness. For different SAR datasets and usage scenarios, feature extraction algorithms need to be redesigned.

In recent years, with the development of deep learning technology, especially the development of convolutional neural networks, the research of image recognition technology has made remarkable progress. The proposal and successful application of AlexNet [13] marked the beginning of the development of deep learning, followed by VGGNet [14] and Restnet [15], which made breakthroughs in natural image recognition. Inspired by the achievements of deep learning in optical images, researchers try to apply deep learning to SAR image recognition. Deep learning is an end-to-end learning method that unifies feature extraction and classification recognition. It can automatically learn the required features based on the learning objectives without the need for additional feature extraction algorithms, reducing manual work and greatly improving the robustness of the algorithm. However, deep learning algorithms require a large amount of training data, while obtaining SAR image training data is difficult compared to optical images. As a result, the available samples for SAR image training are relatively few. Directly applying ordinary optical image recognition algorithms to SAR image recognition can easily cause the overfitting phenomenon [16]–[18]. Therefore, multiple researchers have conducted in-depth research on the application of deep learning in SAR image recognition. Y. Li [19] used CNN networks to extract SAR image features, and used meta-learning training methods and distance metric loss functions to classify images. High recognition accuracy was achieved on both OPENSARSHIP and MSTAR datasets. Jian Guan [20] proposed a CNN network that combines multiple-size convolutional kernels and dense residual networks. The method combines the cosine loss function and cross-entropy loss function to train the network in two stages and performs well in small sample data scenarios. Ying Zhang [21] proposed a training method for convolutional networks, which combines deep metric learning (DML) and an imbalanced sampling strategy to improve classification performance in the imbalanced training sample scenario. Zhang Ting [12] used CNN networks to extract multi-layer depth features of SAR images to improve recog-

Manuscript received November 11, 2023.
Manuscript revised February 28, 2024.
Manuscript publicized August 22, 2024.

[†]Faculty of Electronic Engineering, Naval University of Engineering, Wuhan, P.R. China.
a) E-mail: 192019018@nue.edu.cn (Corresponding author)

DOI: 10.23919/transcom.2023EBP3179

nition accuracy. S. Chen [17] used a fully convolutional network to reduce parameters while achieving high recognition accuracy.

Through the in-depth study of convolutional neural networks, the above algorithms have achieved good results in SAR image recognition scenes. Inspired by the above research, this article proposes a SAR image recognition algorithm based on distance measurement and small-size convolutional networks, which aims to simplify the network structure, prevent overfitting and improve the recognition accuracy. The paper carries out work from optimizing several aspects including loss function, network structure and training method. The main innovations of this paper are as follows:

1. A small-size convolutional kernel convolutional network was constructed to form the backbone of the whole model, which can reduce the number of parameters of the model and improve the final image recognition rate.
2. A loss function based on distance measurement is proposed, which considers the distance from samples to the class centre, the distance between class centres, and the class variance, to guide the model to train in the direction of intra-class aggregation and inter-class dispersion.
3. A convolutional network classifier has been constructed, which reduces model parameters and improves model performance when compared with traditional linear layer classifiers.
4. A two-stage training method is proposed. In the first stage, the distance metric loss function is used to train the feature extraction network, and the cross-entropy is used to train the classification in the second stage. Comparative experiments show the effectiveness of this method.

The rest of this paper is organized as follows: Section 2 introduces the algorithm proposed in this paper, Sect. 3 introduces the experimental results, and Sect. 4 summarises the work.

2. SAR Image Recognition Algorithm Based on Distance Measurement and Small-Size Convolutional Network

2.1 Overall Framework

The overall framework of the image recognition method is shown in Fig. 1. The entire process is divided into three stages. The first stage is the feature extractor training stage, in which the distance measurement loss function is used. After the first stage is completed, it goes into the second stage, in which a convolutional layer is added to the feature extraction network as a classifier and the cross-entropy loss function is used to fine-tune the whole network. After completion, a trained image recognition model is obtained. In the third stage, the performance of the model is tested using a test dataset.

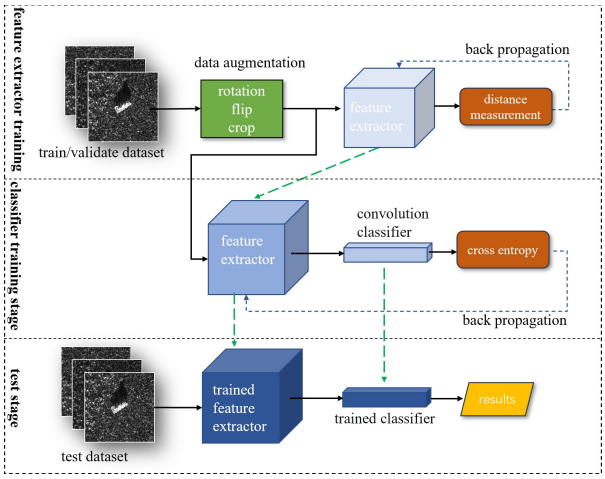


Fig. 1 Overall framework of algorithm.

2.2 Loss Function

In this paper, the algorithm is trained in two stages. In the first stage, the distance measurement method is used to calculate the loss, and in the second stage, the cross-entropy is used to calculate the loss. The loss function is written as follows:

$$loss = \begin{cases} loss_{distance} & 0 < E < E_1 \\ loss_{cross\_entropy} & E_1 < E < E_2 \end{cases} \tag{1}$$

in which E is the epoch variant, E_1 is the epoch num in the first stage, E_2 is the total epoch num.

2.2.1 Loss Function in the First Stage

The formula designed in this paper synthesizes three considerations: the distance from samples to their class centre, the distance between class centres, and the class variance, so as to guide the model to train towards the direction of intra-class aggregation and inter-class dispersion. The formula for the distance measurement loss is:

$$loss_{distance} = \alpha \cdot loss_{s-c} + \beta \cdot loss_{var} + (1-\alpha-\beta) \cdot loss_{c-c} \tag{2}$$

in which, $loss_{s-c}$ is the loss of distance from the sample to its class centre, $loss_{\text{var}}$ is the loss of the class variance, $loss_{c-c}$ is the loss of distance between class centres, α and β are the hyper-parameters. The following describes the detailed calculation methods of the three parts.

$loss_{s-c}$ is calculated by taking the average of the distance loss from each sample to its class centre. The formula is:

$$loss_{s-c} = \frac{\sum_{i=1}^{N} loss(x_i)}{N} \tag{3}$$

where $loss(x_i)$ is the distance loss from the i-th sample to its

class centre, N is the total number of samples in the current batch, $loss(x_i)$ is calculated by the formula:

$$loss(x_i) = -\log\left(\frac{e^{-Dis(x_i,c_{y_i})}}{\sum_{l=0}^{L-1} e^{-Dis(x_i,c_l)}}\right) \tag{4}$$

in which, $y_i \in \{0...L-1\}$ is the label of x_i, and $Dis(x_i, c_l)$is the distance from the i-th sample to the l-th class centre. The paper uses Euclidean distance which is calculated by:

$$Dis(x_i, c_l) = \sqrt{\sum_{m=0}^{M-1} (f_\theta(x_i)_{[m]} - c_{l[m]})^2} \tag{5}$$

Where $f_\theta(x_i)$ is the feature vector output of the sample x_i through the feature extraction network, M is the feature vector dimension and $f_\theta(x_i)_{[m]}$ is the m-th dimension of $f_\theta(x_i)$. c_l is the l-th class centre, which is the mean value of feature vectors belonging to the l-th class in this batch, $c_{l[m]}$ is the m-th dimension of c_l. c_l is calculated by:

$$c_l = \frac{1}{|I(l)|} \sum_{i \in I(l)} f_\theta(x_i) \tag{6}$$

in which, $I \equiv \{1...N\}$ is the set of indices of all input samples in current batch, $I(l) \equiv \{p \in I : y_p = l\}$ is the set of indices of samples whose label is l. $|I(l)|$ is its cardinality. It should be noted that c_l will be updated in each batch.

$loss_{\text{var}}$ is the loss of sample variance in the same class, and is calculated by:

$$loss_{var} = \frac{1}{L} \sum_{l=0}^{L-1} \text{var}_l \tag{7}$$

where var_l is the variance of the l-th class, and is calculated by:

$$\text{var}_l = \sum_{m=0}^{M-1} \text{var}(f_\theta(X(l))_{[m]}) \tag{8}$$

where $X(l) \equiv \{x_i : i \in I(l)\}$ is the sample set whose label is l, $var(.)$ is the variance function.

$loss_{c-c}$ is the loss of distance between class centres. After training, the larger the distance of inter-class, the better the classification effect of the model. However, the absolute distance cannot reflect the distinguishing ability between classes. In this paper, the distance between class centres is compared with the variance of samples belonging to this class. The larger the value, the bigger the difference exists between classes. Figure 2 shows the meaning. Figure 2(a) and Fig. 2(b) have the same distance between class centres, but Fig. 2(a) has a better distinguishing effect. $loss_{c-c}$ is calculated by the formula:

$$loss_{c-c} = \sum_{i=0}^{L-1} \frac{1}{DisFunc(c_i)} \tag{9}$$

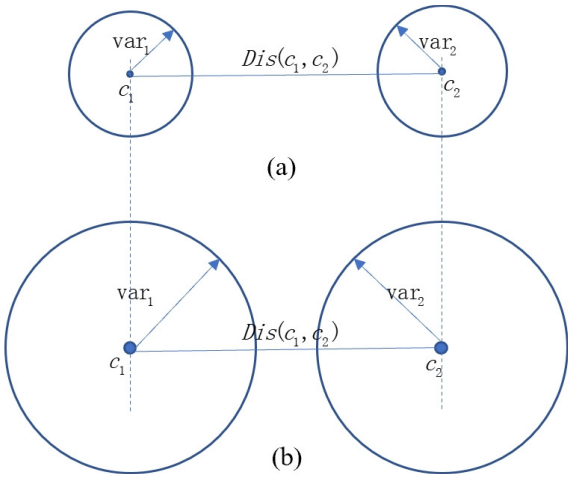


Fig. 2 Influence of the ratio of the distance between category centres and variance on discrimination ability. The distance between the two class centres is the same in (a) and (b), but in (a), the variance is smaller and the discrimination ability is stronger, and in (b), the variance is larger and the discrimination ability is weaker.

$$DisFunc(c_i) = \sum_{l=0}^{L-1} \frac{\text{Dis}(c_i, c_l)}{\text{var}_i + \text{var}_l} \tag{10}$$

in which the l-th class centre c_l is calculated by formula (6), the l-th class sample variance var_l is calculated by formula (8), $Dis(c_i, c_l)$ is the Euclidean distance between the centre of the i-th class and the l-th class.

Distance measurement has been applied to both prototype networks [22], supervised contrast learning [23] and Improved Triplet Loss [24], but there are differences in its connotation. In the prototype network, the distance metric loss function is defined as:

$$loss(x_i) = -\log\left(\frac{e^{-Dis(x_i,c_{y_i})}}{\sum_{l=0}^{L-1} e^{-Dis(x_i,c_l)}}\right) \tag{11}$$

$$loss = \frac{1}{N} \sum_{i=1}^{N} loss(x_i) \tag{12}$$

The loss function of the supervised contrast learning is defined as:

$$\begin{aligned} loss &= \sum_{i \in I} loss(x_i) \\ &= \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log\left(\frac{e^{f_\theta(x_i) \cdot f_\theta(x_p)/\tau}}{\sum_{a \in A(i)} e^{f_\theta(x_i) \cdot f_\theta(x_a)/\tau}}\right) \end{aligned} \tag{13}$$

in which, $I \equiv \{1...N\}$ is the set of indices of all input samples in the current batch, $A(i) \equiv I \backslash \{i\}$ is the set of I excluded i, $P(i) \equiv \{p \in A(i) : y_p = y_i\}$ is the set of indices of all positives in the batch distinct from i, and $|P(i)|$ is its cardinality, τ is the temperature hyper-parameter.

In the Improved Triplet Loss, denote $X_i = < X_i^o, X_i^+, X_i^- >$ as a group of input, in which X_i^o and X_i^+ belong to the same class, X_i^o and X_i^- are in different classes.

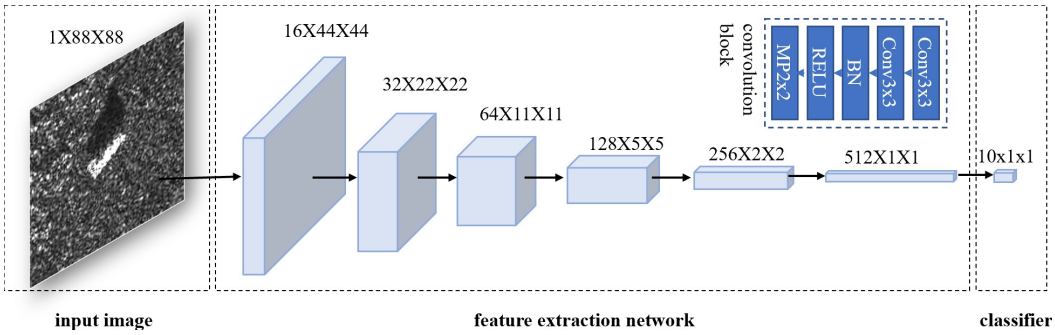


Fig. 3 Network structure of proposed algorithm.

The loss function is defined as:

$$loss = \frac{1}{N}\sum_{i=1}^{N}(\max\{d^n(X_i^o, X_i^+, X_i^-), \tau_1\} + \beta \max\{d^p(X_i^o, X_i^+), \tau_2\})$$

where τ_1, τ_2 and β are the hyper-parameters, and:

$$d^n(X_i^o, X_i^+, X_i^-) = d(f_\theta(X_i^o), f_\theta(X_i^+)) - d(f_\theta(X_i^o), f_\theta(X_i^-))$$
$$d^p(X_i^o, X_i^+) = d(f_\theta(X_i^o), f_\theta(X_i^+))$$
$$d(f_\theta(X_i^o), f_\theta(X_i^+)) = ||f_\theta(X_i^o) - f_\theta(X_i^+)||^2$$

From the above formulas, it can be seen that in the prototype network, only the distance between the sample and the centre of samples in the same class (prototype) is used. In supervised comparative learning, only the distance between the sample and other samples in the same class is considered. In Improved Triplet Loss, only the distance between samples belongs to the same class and different classes are used. So all of them only consider the relationship between the sample and other samples. In addition to the above, this article further considers the internal variance of samples of the same class and the distance between different class centres. This paper simultaneously considers both individual and overall loss of the sample, to guide the model to be trained towards the direction of intra-class aggregation and inter-class dispersion comprehensively.

2.2.2 Loss Function in the Second Stage

The cross-entropy function is used to calculate the loss in the second stage, the formula is:

$$loss_{cross\_entropy} = -\frac{1}{N}\sum_{i=1}^{N}\log(\frac{e^{g(x_i)_{[y_i]}}}{\sum_{l=0}^{L-1} e^{g(x_i)_{[l]}}}) \quad (14)$$

where N is the number of samples in this batch, and, $g(x_i)$ is the output of classifier for the sample x_i, whose dimension is L, $g(x_i)_{[l]}$ is the l-th dimension of $g(x_i)$. $y_i \in \{0...L-1\}$ is the label of x_i, $g(x_i)_{[y_i]}$ is the y_i-th dimension of $g(x_i)$.

2.3 Feature Extraction Network and Classifier

In 2014, Karen Simonyan from Oxford University proposed VGGNet [14] and explored the depth of the network. Due to its simplicity and practicality, it quickly became the most popular convolutional neural network at that time. The inspirations brought by the VGGNet include: (1) Replacing a large convolutional layer with multiple small convolutional layers can obtain the same receptive field size, but significantly reduces parameter size and computational complexity; (2) Using a unified 2x2 max-pool with a stride of 2 to increase local information diversity and reduce feature size, can better capture local information changes and describe edge and texture structures.

Inspired by VGGNet, a feature extraction network composed of small-size convolutional layers and a classifier composed of a convolutional layer is designed, as shown in Fig. 3. The feature extraction network consists of 5 convolutional blocks. Each block is composed of two 3x3 convolutional layers, one Batch-Normalize layer, one RELU activation layer, and one max-pool pooling layer. The pooling layer's size is 2x2 and the stride is 2, it can reduce the feature map size by half. Except for the first convolution block, the number of channels in the first convolution layer of each other convolution block is doubled. In order to reduce network parameters, increase network robustness and avoid overfitting, the 1x1 convolution layer rather than the conventional linear layer is used as the network classifier.

3. Experiments

3.1 Dataset and Parameter Configuration

This article uses the MSTAR dataset to verify the performance of the proposed algorithm. The MSTAR dataset is a public dataset for SAR automatic target recognition provided by the US Advanced Research Projects Agency and the Air Force Laboratory (DARPA/AFRL). The images are obtained by an X-band HH polarized constrained radar with resolution of 0.3 m × 0.3 m. In this paper, 10 types of military ground target data in the Standard Operating Conditions (SOC) of the MSTAR dataset are selected for experiments. The distribution of the dataset is shown in Table 1.

In the experiment, Pytorch was used under Ubuntu 20.04 LTS and GPU RTX 3090 was used to accelerate calculation. The experiment parameters were configured as follows: Batch size (N in Eq. (3)) was 100; The dimension of

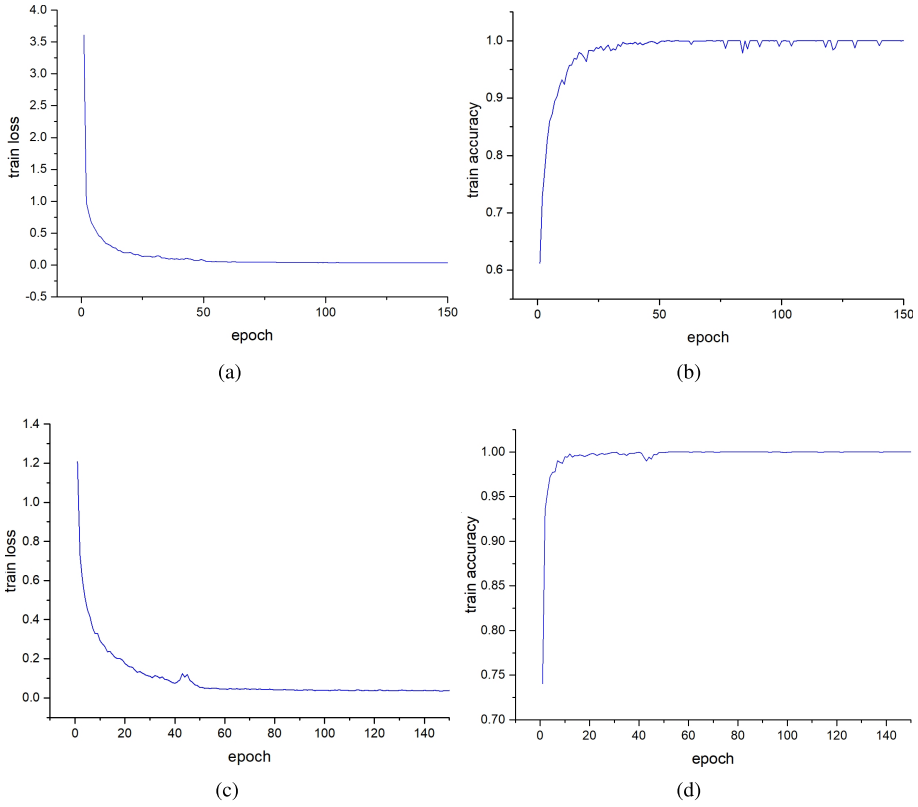


Fig. 4 Model training data (for training data). (a) Loss at stage 1. (b) Accuracy at stage 1. (c) Loss at stage 2. (d) Accuracy at stage 2.

Table 1 MSTAR SOC training dataset and test dataset.

	Training dataset		Test dataset	
Class	**Depression**	**Number**	**Depression**	**Number**
2S1	17°	299	15°	274
BMP2	17°	233	15°	196
BRDM-2	17°	298	15°	274
BTR-70	17°	233	15°	196
BTR-60	17°	256	15°	195
D7	17°	299	15°	274
T62	17°	299	15°	273
T-72	17°	232	15°	196
ZIL131	17°	299	15°	274
ZSU234	17°	299	15°	274

the feature extractor output (M in Eq. (5)) was 512; Adamw optimizer was selected with the learning rate setting to 0.001 and the weight decay setting to 0.004.

3.2 Training and Test Experiments

Figure 4 shows the loss and accuracy changing as the iteration progresses. Figure 4(a) and Fig. 4(b) represent the results in the first stage, and Fig. 4(c) and Fig. 4(d) represent the results in the second stage. As shown in the figure, in the initial stage of training, the loss value drops rapidly and the model converges quickly. In the second stage, the accuracy of training data rises faster than that in the first stage, and the model converges faster. By the 10th round of training, the accuracy has already reached its maximum value, indicating that the model has been adjusted to near the optimal value after the first stage of training.

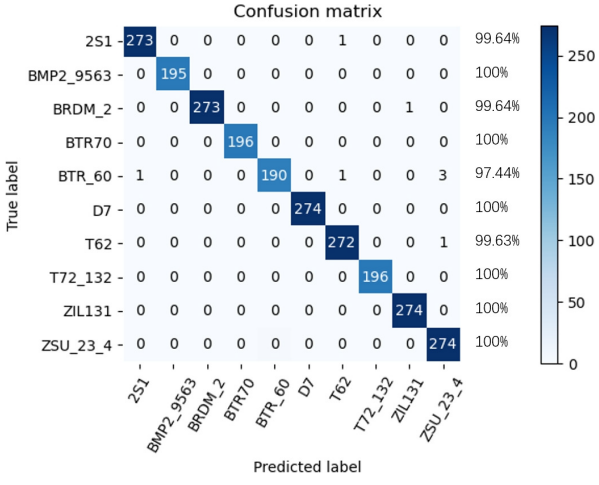


Fig. 5 Confusion matrix.

Figure 5 shows the confusion matrix evaluated by the trained model on the test dataset. It can be seen that the recognition accuracy is high, even reaching 100% for some classes, and the average recognition accuracy is 99.67%, which proves the effectiveness of this model.

3.3 Comparative Experiments

3.3.1 Comparison with Algorithms in Other Literature

In order to further prove the effectiveness of the proposed algorithm, this paper compares the proposed algorithms with literatures [17], [20], [24], classical image recognition convolutional networks VGG16, ResNet-50, and ResNet-18. This article selects data from the MSTAR dataset both in standard operating conditions (SOC) and extended operating conditions (EOC) for comparative experiments. Compared to SOC, EOC data has a greater difference between the training and testing datasets. This article selects two types of EOC datasets (denoted by EOC-1 and EOC-2) that are consistent with reference [17]. In EOC-1, there is a significant difference in the depression angle between the training and test dataset. The training dataset is composed of four targets (2S1, BRDM-2, T-72, and ZSU-234) in the 17° depression angle chosen from Table 1, and the test dataset (shown in Table 2) is composed of data in the 30° depression angle. In EOC-2, there are significant differences in the serial numbers and configurations of targets between the training and test datasets. The training set is composed of four targets (BMP-2, BRDM-2, BTR-70, and T-72) in the 17° depression angle chosen from Table 1, while the test set contains two groups listed in Table 3 (EOC-2-1) and Table 4 (EOC-2-2), corresponding to configuration variants and version variants.

The comparative experimental results are shown in Table 5, in which the results of the literature [17], [20] are obtained from the original data of the literature. Results of VGG16, ResNet-50 and ResNet-18 are obtained through the experiment using the same parameter as the proposed algorithm in this paper. The results of Improved Triplet Loss algorithmn are obtained by replacing the distance measurement method proposed in this paper with that in reference [24], and keeping other parameters and training methods in accordance with this paper. It can be seen from the table that VGG16 performs the worst, followed by ResNet-50. This is mainly due to that the VGG16 and ResNet-50 have a large number of parameters by adopting multiple fully connected layers and multiple layers respectively, and the algorithm complexity comparison can be found in Sect. 3.3.4. Because of the difficulty of obtaining SAR images, the available samples for SAR image training are relatively few, and a model with a big parameter number can easily cause overfitting and performance degradation phenomenon [17]. The models of ResNet-18, literature [17], literature [20], Improved Triplet Loss [24], and this paper have fewer parameters compared to the first two, so their performance is better, reaching over 99% in the SOC dataset. The algorithm proposed in this paper not only achieves optimal performance under the SOC dataset, but also outperforms other algorithms under various EOC datasets, further proving the superiority of the proposed method.

Table 2 EOC-1 test dataset (large depression variant).

Class	Depression	Number
2S1	30°	288
BRDM-2	30°	287
T-72	30°	288
ZSU234	30°	288

Table 3 EOC-2-1 test dataset (configuration variants).

Class	Serial No.	Depression	Number
T-72	S7	15°,17°	419
	A32	15°,17°	572
	A62	15°,17°	573
	A63	15°,17°	573
	A64	15°,17°	573

Table 4 EOC-2-2 test dataset (version variants).

Class	Serial No.	Depression	Number
BMP-2	9566	15°,17°	428
	c21	15°,17°	429
T-72	812	15°,17°	426
	A04	15°,17°	573
	A05	15°,17°	573
	A07	15°,17°	573
	A10	15°,17°	567

3.3.2 Comparison of Different Feature Extraction Networks (Backbones)

To prove the superiority of the feature extraction network of the proposed algorithm, the feature extraction network was replaced as EfficientNetV2 [25] and MobileNetV3 [26], and comparative experiments were conducted on three types of networks under the same condition. The comparison results are shown in Table 6. It can be seen that the feature extraction network used in this paper has a significant advantage in accuracy compared to the other two.

3.3.3 Comparison of Different Distance Measurement Methods

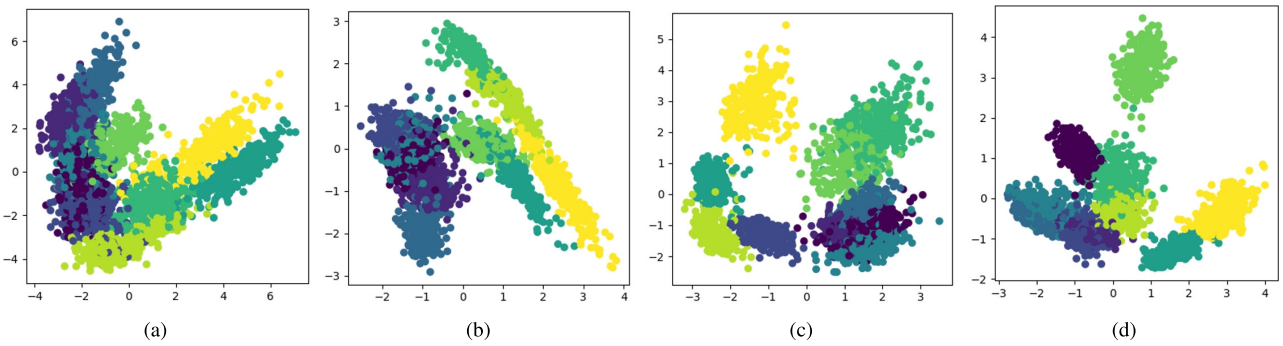
In order to prove the superiority of the distance measurement method proposed in this paper, the loss function trained in the first stage is replaced by the loss function of the prototype

Table 5 Comparison of accuracy with algorithms in other literature.

Algorithm	SOC	EOC-1	EOC-2-1	EOC-2-2
VGG16	96.37%	60.30%	95.31%	96.25%
ResNet-50	98.27%	91.49%	97.20%	98.76%
ResNet-18	99.01%	95.57%	98.63%	99.20%
literature [17]	99.13%	96.12%	98.93%	98.6%
literature [20]	99.26%	-	-	-
Improved Triplet Loss[24]	99.46%	95.13%	97.34%	99.27%
this paper	**99.67%**	**97.18%**	**99.26%**	**99.33%**

Table 6 Comparison of experimental results of different feature extraction networks.

Feature extraction network	Accuracy
EfficientNetV2	97.90%
MobileNetV3	98.52%
this paper	**99.67%**



(a) (b) (c) (d)

Fig. 6 Visualization of the output features of models trained with different loss functions after PCA reduction. (a) Prototype network loss function. (b) Supervised contrast learning network loss function. (c) Improved Triplet Loss. (d) This paper.

Table 7 Comparison of different distance measurement loss functions.

Loss function	Accuracy
prototype network[22]	99.55%
supervised contrast learning[23]	99.59%
Improved Triplet Loss[24]	99.46%
this paper	**99.67%**

Table 8 Comparison of algorithm complexity.

Algorithm	Parameter size	FLOPs	Accuracy
VGG16	134.3004M	2,442,049,536	96.37%
ResNet-50	23.5222M	685,958,400	98.27%
ResNet-18	11.1754M	292,672,256	99.01%
EfficientNetV2_S	20.1899M	484,495,360	97.90%
MobileNetV3_S	2.5426M	**11,515,184**	98.52%
this paper	**1.7757M**	111,825,172	**99.67%**

network [22], the supervised contrastive learning network [23] and Improved Triplet Loss [24]. Experiments were conducted under the same conditions, and the comparison results are shown in Table 7. It can be seen that the algorithm proposed in this paper performs best. Using three different loss functions to train the model, the output features are dimensionally reduced by PCA, and the visualization results are shown in Fig. 6. It can be seen that the algorithm in this paper distinguishes each category more clearly and has a stronger classification ability.

3.3.4 Comparison of Algorithm Complexity

Table 8 provides a comparison between our algorithm and other algorithms in terms of parameter size, floating point operations (FLOPs), and accuracy. It can be seen that the algorithm in this paper has the minimum parameter size, and FLOPs are only higher than MobileNetV3_S, but the accuracy is the highest. The overall algorithm complexity is low while ensuring a high recognition rate.

3.4 Ablation Experiment

In order to verify the effectiveness of the three design ideas proposed in this paper, namely, convolution classifier, small-size convolution kernel and distance measurement loss function(represented by A, B, and C respectively), this paper uses several variants of the algorithm to conduct ablation experiments, namely Variant 1: use a linear layer as the final classifier; Variant 2: use a 7x7 convolutional kernel instead of three 3x3 convolutional kernels, and use a 5x5 convolutional kernel instead of two 3x3 convolutional kernels; Variant 3: do not use the distance measurement loss function, and directly use cross entropy for training. The experimental results of the three variants are shown in Table 9, which shows that all three variants perform varying degrees of reduction in accuracy. Figure 7 shows their training processes, and it can be seen that after using the two-stage training method, the convergence speed of the second training is significantly accelerated due to the completion of the first training. From the embedded zoomed figure, we can see that Variant 1 has the worst performance, followed by Variant 2 which has

Table 9 Ablation experiment result of variant algorithms.

Algorithm	A	B	C	Accuracy
Variant 1: linear classifier		✓	✓	98.56%
Variant 2: big convolution kernel size	✓		✓	99.38%
Variant 3: don't use distance measurement loss	✓	✓		99.55%
No change	✓	✓	✓	**99.67%**

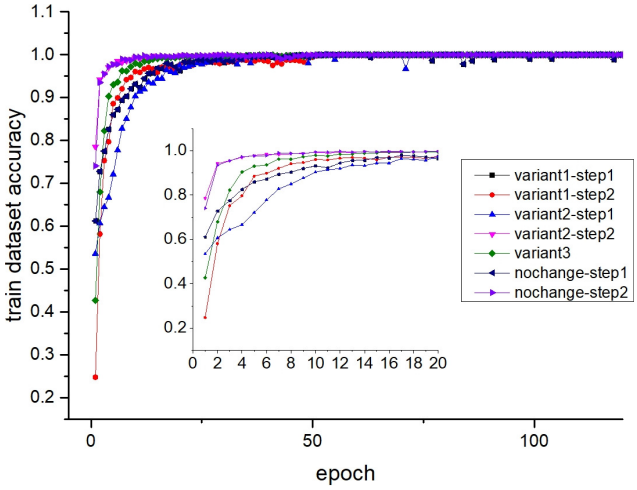
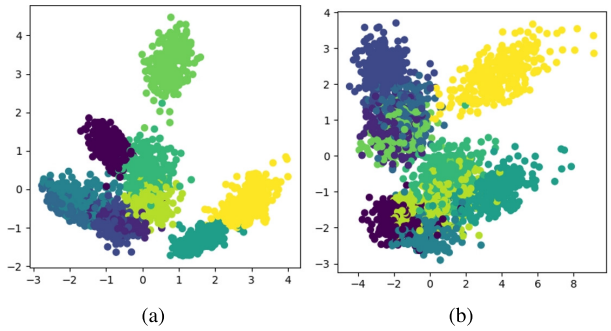


Fig. 7 Comparison of the training process of variant algorithms.



(a) (b)

Fig. 8 Visualization of feature extraction network's output after PCA dimensionality reduction. (a) The model trained with distance measurement loss. (b) The model trained without distance measurement loss.

large-scale convolution kernels, followed by Variant 3 with the model that does not use the distance metric. The model that does not make any changes has the best performance. This proves the effectiveness of convolution classifier, distance measurement loss function and small-size convolution. In addition, from the experimental results, it can be seen that the accuracy of variant 1 and variant 2 are relatively low, which can be ascribed to that the use of linear classifiers and large convolutional kernels has increased the number of model parameters to a certain extent, and the overfitting occurs. The phenomenon of performance degradation caused by increasing the number of parameters is also reflected in reference [17]. In this paper, the model can reduce the number of parameters by simplifying the network, thus supress overfitting to a certain extent.

Figure 8 shows the visualization of the feature extraction network's output after PCA dimensionality reduction. Figure 8(a) and Fig. 8(b) show the results of the trained model with and without distance measurement loss function, respectively. It can be seen that after training with the distance measurement loss function, categories are distinguished more clearly and the classification ability is stronger.

3.5 Hyper-Parameter Settings

α and β in Eq. (2) are the two important parameters of the loss function in this paper. Experiments were conducted by changing α and β from 0.1 to 0.8 (we choose the value meet $\alpha+\beta < 1$ since the ratio of $loss_{c-c}$ is 1-α-β) respectively, other parameters are consistent with that in Sect. 3.1, and the accuracy results are shown in Fig. 9. It can be seen that high accuracy can be obtained at several points such as (α,β)={(0.8,0.1),(0.3,0.2),(0.7,0.2)}. The paper chooses the point of (α=0.8, β=0.1).

4. Conclusion

In order to improve the accuracy of SAR image recognition, this paper proposes a small-size full convolutional network based on distance measurement. The feature extraction part of the network is composed of multiple 3x3 convolution layers, and the classifier of the network is composed of a 1x1

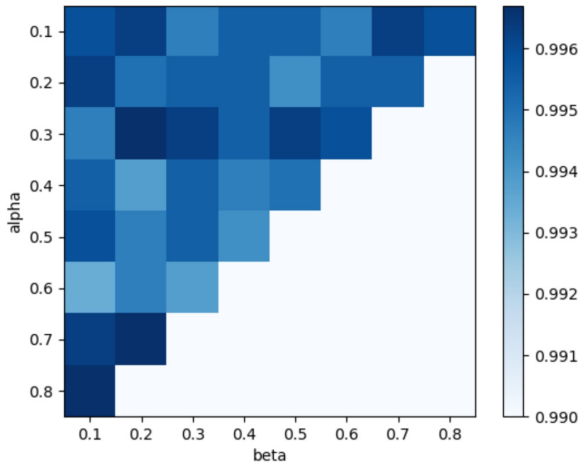


Fig. 9 Results with different α and β settings.

convolution layer. The design of small-size and convolution classifiers improves accuracy while reducing network parameters and computational complexity. A loss function based on distance measurement is designed, which makes comprehensive use of the distance from the sample to the category centre, the distance between category centres and the variance of samples in the same category. Feature map visualization shows, after training with the distance measurement loss function, categories are distinguished more clearly and the features of samples in the same category are more clustered, the classification ability is stronger. Finally, multiple comparative experiments have shown that the method proposed in this paper is superior to the methods proposed in other papers and classic image recognition models.

Acknowledgments

The authors would like to thank Jinling Xing for full text statement and logicall review, and thank their colleagues in the lab for suggestions.

References

[1] R. Yang, Z. Hu, Y. Liu, and Z. Xu, "A novel polarimetric sar classification method integrating pixel-based and patch-based classification," IEEE Geosci. Remote Sens. Lett., vol.17, no.3, pp.431–435, 2019.

[2] C. Cao, Z. Cui, L. Wang, J. Wang, and J. Yang, "Cost-sensitive awareness-based SAR automatic target recognition for imbalanced data," IEEE Trans. Geosci. Remote Sens., vol.60, pp.1–16, 2022.

[3] Y. Zilu, X. Chen, Z. Yikui, and W. Faguan, "Self-attention multiscale feature fusion network for small sample SAR image recognition," Journal of Signal Processing, vol.36, no.11, pp.1846–1858, 2020.

[4] R. Shang, J. Wang, L. Jiao, S. Rustam, B. Hou, and Y. Li, "SAR targets classification based on deep memory convolution neural networks and transfer parameters," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol.11, no.8, pp.2834–2846, 2018.

[5] C. Cao, Z. Cui, J. Wang, Z. Cao, and J. Yang, "A demand-driven SAR target sample generation method for imbalanced data learning," IEEE Trans. Geosci. Remote Sens., vol.60, p.15, 2022.

[6] U. Srinivas, V. Monga, and R.G. Raj, "SAR automatic target recognition using discriminative graphical models," IEEE Trans. Aerosp.

Electron. Syst., vol.50, no.1, pp.591–606, 2014.
[7] C. Belloni, A. Balleri, N. Aouf, J. Caillec, and T. Merlet, "Explainability of deep SAR ATR through feature analysis," IEEE Trans. Aerosp. Electron. Syst., vol.57, no.1, pp.659–673, 2021.
[8] S. Feng, K. Ji, X. Ma, L. Zhang, and G. Kuang, "Target region segmentation in SAR vehicle chip image with ACM net," IEEE Geosci. Remote Sens. Lett., vol.19, pp.1–5, 2022.
[9] M. Amoon and G.A. Rezai-Rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of zernike moments features," IET Computer Vision, vol.8, no.2, pp.77–85, 2013.
[10] H. Yan, B.Y. Ping, and Z.X. Fei, "Synthetic aperture radar target recognition based on KNN," Fire Control & Command Control, vol.43, no.09, pp.111–113+118, 2018.
[11] T. Wu, J. Xia, and Y. Huang, "Target recognition method of SAR images based on cascade decision fusion of SVM and SRC," Journal of Henan Polytechnic University (Natural Science), vol.39, no.04, pp.118–124, 2020.
[12] Z. Ting and C. De-Rao, "SAR target recognition based on joint use of multi-level deep features," Fire Control & Command Control, vol.45, no.02, pp.135–140, 2020.
[13] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol.25, no.2, 2012.
[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.770–778, 2016.
[16] S. Wang, Y. Wang, H. Liu, and Y. Sun, "Attribute-guided multi-scale prototypical network for few-shot SAR target classification," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol.14, pp.12224–12245, 2021.
[17] S. Chen, H. Wang, F. Xu, and Y.Q. Jin, "Target classification using the deep convolutional networks for SAR images," IEEE Trans. Geosci. Remote Sens., vol.54, no.8, pp.4806–4817, 2016.
[18] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for SAR target classification with limited labeled training data," IEEE Geosci. Remote Sens. Lett., vol.14, no.7, pp.1091–1095, 2017.
[19] Y. Li, X. Li, Q. Sun, and Q. Dong, "SAR image classification using CNN embeddings and metric learning," IEEE Geosci. Remote Sens. Lett., vol.19, pp.1–5, 2022.
[20] J. Guan, J. Liu, P. Feng, and W. Wang, "Multiscale deep neural network with two-stage loss for SAR target recognition with small training set," IEEE Geosci. Remote Sensing Lett., vol.19, pp.1–5, 2022.
[21] Y. Zhang, Z. Lei, H. Yu, and L. Zhuang, "Imbalanced high-resolution SAR ship recognition method based on a lightweight CNN," IEEE Geosci. Remote Sens. Lett., vol.19, pp.1–5, 2022.
[22] J. Snell, K. Swersky, and R.S. Zemel, "Prototypical networks for few-shot learning," Advances in Neural Information Processing Systems, vol.30, pp.4077–4087, 2017.
[23] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," Advances in Neural Information Processing Systems, vol.33, pp.18661–18673, 2020.
[24] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.1335–1344, 2016.
[25] M. Tan and Q.V. Le, "EfficientNetV2: Smaller models and faster training," International Conference on Machine Learning, pp.10096–10106, PMLR, 2021.
[26] A. Howard, M. Sandler, G. Chu, L.C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, and V. Vasudevan, "Searching for MobileNetV3," Proc. IEEE/CVF International Conference on Computer Vision, pp.1314–1324, 2019.

Yuxiu Liu received the B.S. in Communication engineering from Central South University in 2006, received the M.S. degrees in Signal and information processing from South China University of Technologyin 2013. During 2013–2019, she stayed in Microsoft (China) Co. LTD, studied news-related data. During 2019–2023, she is a lecturer in Naval University of engineering. Her research interests focus on image recognition.

Na Wei born in 1980, received Ph.D. from Beijing University of Aeronautics and Astronautics in signal and information processing, reveived M.S. from Naval University of engineering in computer application technology, His research interests focus on pattern recognition.

Yongjie Li born in 1977, received Ph.D. from Naval University of engineering in systems engineering and M.S. from Naval University of engineering in computer application technology. He is an associate professor in Naval University of engineering. His research interests focus on database and information techniques.