Meta-Learner-Based Stacking Network on Space Target Recognition for ISAR Images

Yun Zhang, *Member, IEEE*, Haoxuan Yuan, *Student Member, IEEE*, Hongbo Li, Jiaying Chen, and Muqun Niu

***Abstract*—Recently, the deep learning models have achieved great success in the recognition of inverse synthetic aperture radar (ISAR) images. However, most of the deep learning models fail to obtain satisfactory results under the condition of small samples due to the contradiction between the large parameter space of the deep learning models and the insufficient labeled samples of space target imaging by ISAR. In this article, a method of meta-learner-based stacking network (MSN) is proposed, which can realize the high-precision classification of space target by ISAR images under the condition of small sample. Innovatively, a rotation-invariant attention mechanism (RAM) module is added into Resnet50 network to magnify the difference of embedded features of target and background. Complementarily, the deep relationship between the features of fine-grained ISAR image is extracted by using graph convolutional network and relation network. Finally, an innovative adaptive weighted XGBoost algorithm is used to integrate the prediction results of the base learners. The main contributions of this article include proposing a RAM module and using an innovative adaptive weighted XGBoost algorithm to realize ensemble learning. The experiment results show that the RAM module effectively concentrates the network's attention on the recognized target, and the recognition rate of MSN is about 5% higher than that of a single base learner under different data volume conditions, which proves that MSN achieves competitive accuracy against other state-of-the-art approaches.**

***Index Terms*—Ensemble learning, graph convolutional network (GCN), inverse synthetic aperture radar (ISAR image), small sample set, space target.**

I. Introduction

INVERSE synthetic aperture radar (ISAR) is a kind of imaging radar with a high resolution in both range and azimuth dimensions [1]. It obtains high-range resolution by transmitting broadband signals [2], and uses the Doppler signal generated by the relative motion between the target and the radar to obtain high-range resolution, which can enable it to perform high-resolution imaging of targets such as manned spacecraft, satellites, and space stations. In this way, ISAR overcomes the disadvantages that aerial cameras, optical video cameras, infrared cameras, and other equipment are limited by climatic conditions [3].

At present, the research of ISAR imaging technology continues to deepen, and the generated ISAR images can reflect the fine structure characteristics of components such as solar panels, engines, and antennas of the target, which is the technical basis for space target classification based on ISAR images [4]. The classification of ISAR images of space targets belongs to fine-grained recognition, which pays more attention to the subtle differences between various satellites. Affected by factors such as measurement conditions and imaging principles, it is difficult to extract the features of space target ISAR images. It is mainly reflected in the following aspects: First, due to the influence of speckle noise, interference fringes, and other factors, the quality of ISAR images will decline and it is difficult to identify types. Second, ISAR imaging methods are sensitive to imaging parameters such as incident angle. There are differences between different individuals of the same type of spacecraft, and there are many possibilities for the spatial distribution of 3-D scatters of space targets in 2-D ISAR projection. Furthermore, variations in structure may lead to changes in the space target itself, and there may be differences in targets of the same type, such as different pivotal payloads on them. For the above reasons, ISAR image recognition of space targets often faces a small sample condition. Finally, ISAR image is high-dimensional, and extracting features directly will lead to low computational efficiency. However, after extracting low-dimensional information, it must be considered whether the useful information is lost. These features bring difficulties to the classification of ISAR images.

In conclusion, robust feature extraction and appropriate dimension compression are the key to determine the final classification effect of ISAR images. The authors in [5]–[11] extract the stable geometric shape features of the ISAR image, and finally obtains a good recognition result. The authors in [12]–[19] project the ISAR target image into another feature space to form a feature vector, whose advantage is that the amount of calculation is reduced and the features are more stable. He *et al.* [20] proposed that local preserving projection (LPP) algorithm can be used for feature extraction and target recognition of 2-D ISAR images. LPP algorithm has the advantages of linear subspace method and flow learning. Therefore, compared with the traditional linear subspace methods such as principal component analysis (PCA) and kernel PCA, it has better recognition effect. However, the algorithm does not consider the influence

Manuscript received June 29, 2021; revised September 7, 2021 and October 29, 2021; accepted November 13, 2021. Date of publication November 18, 2021; date of current version December 8, 2021. This work was supported by the National Natural Science Foundation of China under Grants 61201308 and 61671490. *(Corresponding author: Hongbo Li.)*

Yun Zhang, Haoxuan Yuan, Hongbo Li, and Muqun Niu are with the School of Electronic and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: zhangyunhit@hit.edu.cn; 1037792840@qq.com; drbobo@hit.edu.cn; 544829300@qq.com).

Jiaying Chen is with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: chen98jia@163.com).

Digital Object Identifier 10.1109/JSTARS.2021.3128938

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

of translation, rotation, and scale change of ISAR images on the classification effect. Sang-Hong Park used the polar mapping classifier for classification in [21]. Transferring the ISAR image to the polar coordinate system can convert the rotation of the image into the angular direction in the polar coordinate system. The translation on the above makes it have scale translation and rotation invariance, and then the PCA method is used to compress the image and perform target recognition, and the recognition rate is higher. However, the rotation of the coordinate system may lead to the loss of image information and affect the recognition. Elyounsi *et al.* [22] use edge detection and Fourier descriptor to extract the boundary shape features of ISAR images. This method may lead to limited classification effect because it only uses shape features for classification. At the same time, due to the influence of projection surface, different targets may have very similar shape features. Xu *et al.* [23] use geometric moment invariants for ISAR image recognition, which can make full use of the internal information of target shape. However, this method is sensitive to noise and shape changes. Park *et al.* [24] uses 2-D Fourier transform and polar mapping for ISAR image classification, which has good classification effect under the condition of small amount of data. But this method assumes that the training data and test data are located on the same projection plane, resulting in limited practical application effect.

Although the above methods have a good attempt in ISAR image recognition, they cannot solve the small sample problem of ISAR image recognition of space targets mentioned above. In recent years, meta-learning has developed rapidly and has been used as a common method for small sample recognition [25]. Small sample learning algorithms based on meta-learning mainly include algorithm based on metric [26]–[28], algorithm based on data enhancement [29]–[30], and algorithm based on model optimization [31]–[33]. The small sample learning method based on metric is a simple and effective way. Siamese network is realized by sharing network weights [25]. Siamese network has two inputs, which are mapped to a new space through two neural networks sharing weights, so that the similarity of the two samples can be calculated in this new space and the network finally achieves the purpose of classification. Matching network first uses a convolutional neural network to obtain the shallow representation of support set and test set, then puts them into a bidirectional LSTM network, and finally calculates the cosine similarity of output features to represent the similarity between query set image and support set image [26]. Prototypical network maps an input image to a potential space [27]. Among them, the prototype of a category is obtained from the vectorized sample data of all images of the same category in the support set. Then, by calculating the Euclidean distance between the vectorized value of the query set image and the category prototype, the category of the query set image can be obtained.

Aiming at the difficulties of space target ISAR image recognition proposed above, this article proposes an ISAR image recognition method of space targets based on stacking network under the condition of small sample. Meta-learner-based stacking network (MSN) proposed has three base learners: Resnet50 network with rotation-invariant attention mechanism (RAM) module, graph convolutional network (GCN), and relation network (RN). The last meta-learner adopts an adaptive weighted XGBoost method. MSN can effectively solve the difficulties of the ISAR images recognition of space target proposed above, which is specifically reflected in the following: First, the RAM module is added between the convolution layers of resnet50 to make the network pay more attention to the region where the space target is located in the ISAR image and not sensitive to the irrelevant noise background. In this way, more key feature information in the image can be extracted [34], which cannot only improve the accuracy of image classification, but also improve the stability of the final training result of the model [35]–[39]. Second, the combination of resnet50 and RAM pays more attention to the semantic information of ISAR images of space target. At the same time, the network should also pay more attention to the deep feature relationship between ISAR fine-grained images [40]–[46], so two other base learners are added: GCN and RN. The GCN extracts the deep feature relationship of fine-grained images by constructing the adjacent matrix between nodes [47], and the RN finds the images of the same category by calculating the distance between image features [48]. Third, in the final meta-learning part, an innovative adaptive weighted XGBoost method is used to relearn to obtain the discriminant model by adding a layer of meta-learner and using meta-features as input [49]. It can realize the effective integration between multiple different models and further enhance the learning ability of the model.

To sum up, this article proposes a new ISAR image classification method of space target, which realizes the integration of multiple networks through stacking algorithm. The main contributions of this method are as follows:

1) A RAM module is added into layers of Resnet50, which can allow the network to extract more critical feature information in the fine-grained ISAR image of space target, and weaken the influence of irrelevant background. By paying more attention to semantic information, dependence on a large number of effective data of Resnet50 network is reduced.
2) Through the introduction of GCN and RN, the relationship between the depth features of space target ISAR images of is found. By complementing each other with the semantic information extracted by Resnet50, the extracted deep relationship can further improve the network recognition ability.
3) This article proposes an innovative adaptive weighted XGBoost method by which a meta-learner can be constructed. Adding it at the end of the three base learners can integrate the prediction results of the base learners, improve the intelligent recognition ability of the network, and achieve competitive accuracy against other state-of-the-art approaches.

The rest of this article is organized as follows. In Section II, the difficulties in ISAR image recognition of space targets and preprocessing of ISAR images are described in detail. In Section III, the specific network structure of the method in this article is presented. In Section IV, experiment results and

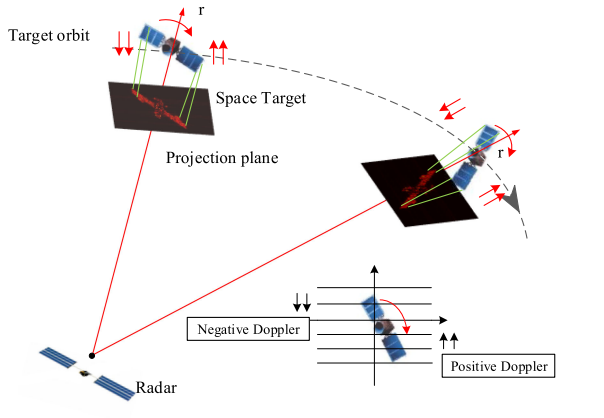


Fig. 1. Geometric relationship between the radar and the target.

detailed analysis are presented. Finally, Section V concludes this article.

II. Image Simulation and Preprocessing

A. Difficulties in ISAR Image Recognition of Space Targets

The task of space target recognition faces many difficulties and challenges. ISAR images of space targets need to have good interpretability in order to be easily recognized, and the features extracted in the recognition process need to be stable and representative. However, because ISAR imaging methods are sensitive to imaging parameters such as incident angle, and there are differences between different individuals of the same type of spacecraft, classification often faces the problem of small samples. The difficulties in ISAR image recognition of space targets are specifically embodied in the following:

First, the change of the motion state of the space target affects the imaging result. Fig. 1 shows the turntable model for ISAR imaging and the geometric relationship between radar and target. The motion of the space target can be equivalent to a translation and a rotation, and the Doppler generated by each scattering point on the target in the translation is exactly the same, which does not contribute to ISAR imaging. Therefore, after the translation is compensated, the target becomes a turntable model. During the rotation of the space target relative to the spaceborne radar, the approaching part produces positive Doppler and the part getting away produces negative Doppler, that is, the result of ISAR imaging depends on the movement state of the target relative to the radar. However, space targets are often noncooperative. In the process of ISAR imaging of space targets by spaceborne satellites, due to the ascent and descent of the target, the ISAR projection surface changes with the change in the rotation direction of the target relative to the imaging satellite. This phenomenon will cause the ISAR image to display different sides of the target, which will affect the recognition of the target.

Second, there is clutter interference caused by complex electromagnetic environment in space, including space debris, occlusion, stacking, and electromagnetic interference, which has a great impact on image recognition. Therefore, the image preprocessing operation must be carried out before the recognition of the ISAR image.

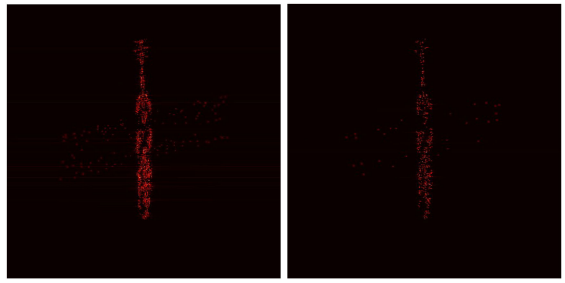


Fig. 2. ISAR images with different resolutions.

Third, the imaging mechanism of ISAR will affect the recognition. The target is sensitive to imaging parameters, including elevation angle, imaging frequency, imaging mode, polarization mode, number of sights, and signal-to-noise ratio. For example, the wider the frequency band of the signal, the

higher the vertical resolution of the image; and the longer the accumulation time of imaging, the higher the horizontal resolution of the image. The simulation of ISAR images in this article uses the Range-Doppler (RD) algorithm, and the specific derivation can be found in [50]. In addition, the angular glint phenomenon of ISAR images has an impact on image recognition. In the ISAR imaging process, the target is considered to be composed of scattered points. As the radar gets closer to the target, random changes in the scattering intensity and relative phase of different parts of the complex target cause flickering points on the image, making the image appear variety. Fig. 2 shows two ISAR images of space target with different resolutions. The angular glint phenomenon on them is obvious, the brightness of the scattering points is obviously different, and this difference changes with the angle of sight line. The resolution of the image will also affect the classification of the images.

Finally, as the function of space targets is very specific, the same type of targets may have different shapes, such as different structures and different pivotal payloads. It is necessary to use some methods that can extract deep features of images, such as GCN, to solve the above problems.

Due to the problems mentioned above, it is urgent to explore robust space target recognition methods under small sample conditions. Fig. 3 shows the overall difficulties in ISAR recognition of space targets.

B. ISAR Image Preprocessing Based on ELF

Each scattering unit in ISAR image can be regarded as a collection of many small scattering points. The echo signal intensity of different scattering units is random, and it will lead to speckle noise in the imaging process and the decline of ISAR image quality, which brings difficulties to ISAR image recognition. Therefore, before target recognition, it is necessary to process the ISAR image and filter the noise. This article

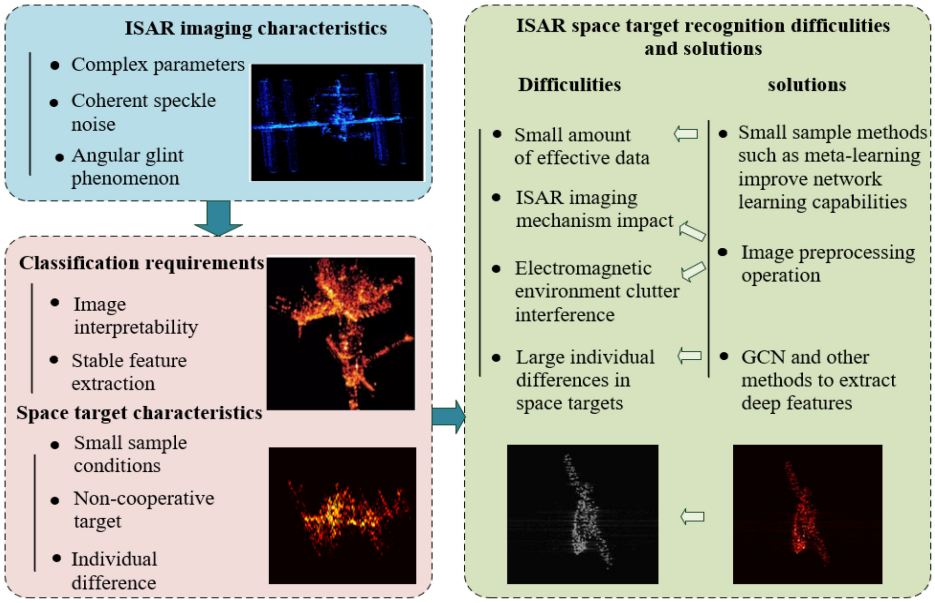


Fig. 3. Overall difficulties in ISAR image recognition.

chooses the enhanced Lee filtering (ELF) method to filter the speckle noise of ISAR images.

Speckle noise in ISAR images is usually considered as a multiplicative noise model, which can be expressed as

$$I(t) = x(t) \times u(t). \tag{1}$$

In formula (1), $I(t)$ is the image after interfered by noise, $x(t)$ is the scattering cross-section of random ground backward target, and $u(t)$ represents noise.

In Lee filter algorithm, the multiplicative noise model is first approximated to a linear system, and then the formula of Lee filter can be obtained by using minimum mean square error (MSE) criterion:

$$w = \overline{I} + k(I - \overline{I}) \tag{2}$$

where $\overline{I}$ is the average pixel intensity of the filter window, w is the estimation of the image affected by noise, and k is the weighting coefficient. The expression of k is

$$k = \frac{\sigma_I^2 - \sigma_n^2 \overline{w}^2}{\sigma_I^2(1 + \sigma_n^2)} \approx 1 - \frac{C_n^2}{C_I^2} \tag{3}$$

where

$$C_n^2 = \frac{\sqrt{\mathrm{var}(n)}}{\overline{n}}, C_I^2 = \frac{\sqrt{\mathrm{var}(I)}}{\overline{I}}.$$

When the speckle noise is not completely generated, the Lee filter cannot achieve a satisfied effect. Therefore, this article chooses to use the ELF method. The expression of above is rewritten as

$$w = \begin{cases} \overline{I}, C_I < C_{\min} \\ \overline{I} + k(I - \overline{I}), C_{\min} < C_I < C_{\max} \\ I, C_I > C_{\max} \end{cases} \tag{4}$$

where C_I represents the local standard deviation coefficient of the image, $C_{\min} = C_n$, $C_{\max} = \sqrt{1 + \frac{2}{L}}$, and L represents the equivalent number of views.

Specifically, the algorithm divides the image regions into three categories, each of which is processed differently as follows:

1) The region of $C_I < C_u$ (u represents noise) represents homogeneous region, and the average value of pixels in sliding window is used as the center value in this region.
2) The region of $C_u < C_I < C_{\max}$ represents a weak texture region, and the traditional Lee filter is used in this region.
3) The region of $C_I > C_{\max}$ represents heterogeneous region, and the original value is directly retained in this region.

Then, power transformation is performed on the image. The pixel value of ISAR image is generally concentrated, which causes the image contrast is not very high. In order to improve the contrast ratio of ISAR image, we introduce the power transformation method. The form of power transformation is as follows:

$$J(i,j) = [H(i,j)]^{\alpha} \tag{5}$$

where changing the value of the parameter α can adjust the contrast ratio of the image and improve the accuracy of the recognition.

After power transformation, energy normalization is performed on the image. Energy normalization can find the invariant in the image through the invariant moment of the image. For ISAR images, the scattering intensity changes with the distance between radar and target, which affects the quality of ISAR image. For ISAR images, operating energy normalization can weaken and eliminate this effect. The formula of energy normalization is as follows:

$$K(i,j) = \frac{J(i,j)}{\sqrt{\sum_i \sum_j |J(i,j)|^2}} \tag{6}$$

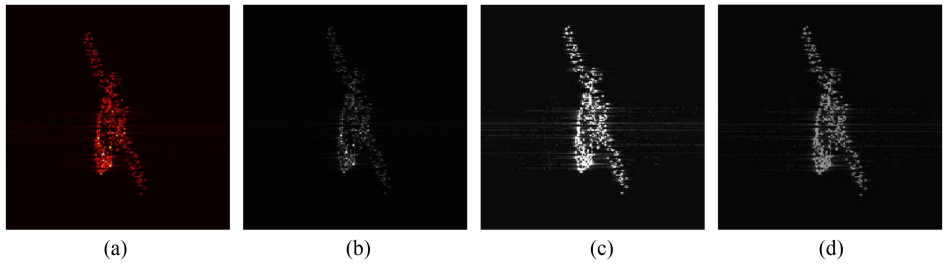


Fig. 4. ISAR image preprocessing example. (a) Original data. (b) Enhanced Lee filter. (c) Power transformation. (d) Energy normalization.

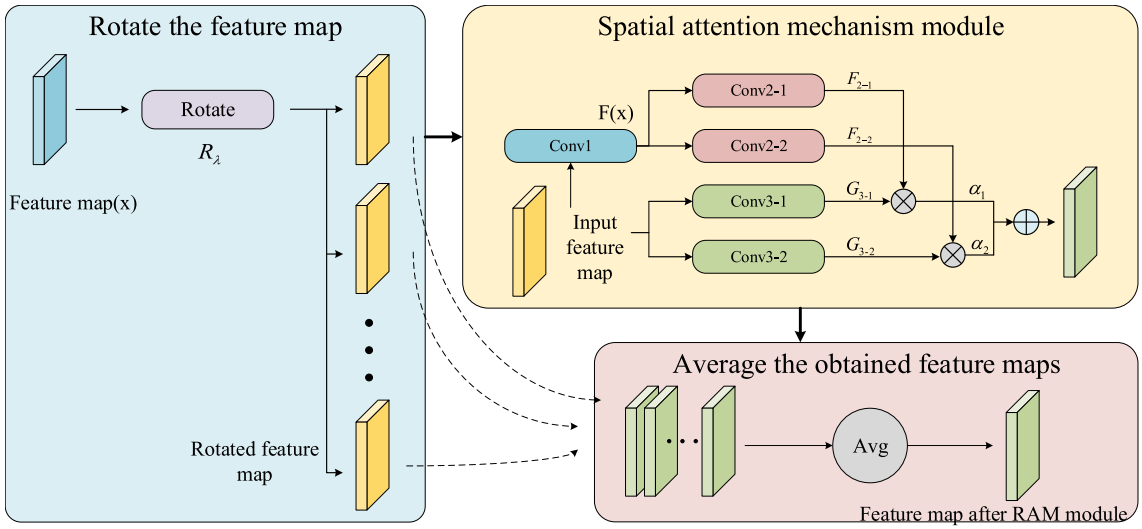


Fig. 5. Structure of RAM module.

where $J(i,j)$ and $K(i,j)$ represent the original image and the normalized image, respectively. Fig. 4 shows an example of ISAR image preprocessing. It can be seen that after the above preprocessing process, the speckle noise of ISAR image becomes less obvious, and the space target is more obvious than that in the original figure, which is convenient for subsequent recognition. Finally, the image dataset is expanded by four times through left-right mirroring and up-down mirroring, making the training results more robust.

III. Network Structure

A. Resnet50 With RAM Module

Resnet50 is a network composed of the residual learning modules. In the training process, the depth error is directly transferred to the shallow layer through the residual module. It contains four residual blocks. Finally, the full-connection operation is performed to facilitate the classification task. As an excellent deep convolution network classification model, Resnet50 needs a large number of effective images. An innovative module called RAM module is introduced after the blocks of Resnet50. This module can let the network extract the key feature information in the image, and weaken the influence of irrelevant background. Aiming at the problem of less labeled sample data, this article constructs an attention mechanism network model for small sample learning, which can improve the ability of existing models to recognize ISAR images. At the same time, the stability of the final training results is improved.

The structure of RAM module is shown in the fig. 5. The input feature map of RAM module is $x \in R\hat{}\{C \times W \times H\}$, where C is the number of channels of the feature map, $W \times H$ is the size of the feature map. The input feature map is rotated by the rotation function R_λ to become a rotated feature map. $R_\lambda = \{R_{\lambda_1}, R_{\lambda_2}, \ldots.., R_{\lambda_K}\}$ is a family of K rotation transformations with R_{λ_i} denoting the rotation operation of an input feature map with the angle of $\lambda_i \in \lambda, i = 1, \ldots, K$. The feature map is represented in the form of a matrix, and its rotation operation is shown in Fig. 6. The left side of the figure is a 4×4 feature map, the upper left corner position is marked as $p(x,y)$, valued as a_{11}. After rotation, the new position is $p'(x,y)$, and its four adjacent points are (x,y), $(x+1,y)$, $(x,y+1)$, and $(x+1,y+1)$. After calculating the Euclidean distances between this point and four adjacent points, the value of the (x,y) point becomes a_{11} by the nearest neighbor interpolation method. If there is an unassigned point, keep the original value.

The rotated feature map is input into Conv1 which is the first convolution module. The convolution kernel size of Conv1 is 3×3, and the number of channels is the number of characteristic channels of the input. Then, a rectified linear unit (ReLU) function is used as activation function for processing, which

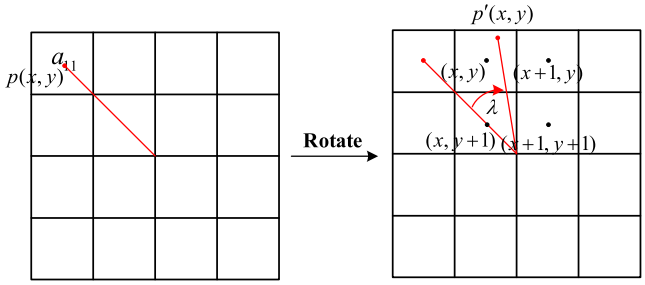


Fig. 6. Schedule of rotation operation.

can be expressed as

$$F(x) = \mathrm{ReLU}(\mathrm{Conv1}(R_\lambda(x))). \tag{7}$$

For Conv2-1 and Conv2-2 module, the input feature map is $F(x)$, the convolution kernel size is 1×1, and the number of channels is the same as the number of input characteristic channels. Next, the obtained feature map is processed by Softmax function to get the attention feature map of the network, F_{2-1} and F_{2-2}. The expression formula of the attention feature map is

$$F_{2-i} = Softmax(\mathrm{Conv}_{2-i}(F(x))), i = 1, 2. \tag{8}$$

Conv3 module uses dilated convolution, and the convolution kernel size is 3×3. Dilation step is set to 2, parameter filling is set to 2, and the number of channels is the same as the number of input characteristic channels. Then, the batch normalization operation is used to get the output feature map, $G_{3-i}, i = 1, 2$, which can be expressed as

$$G_{3-i} = \mathrm{BN}(\mathrm{Conv}_{3-i}(R_\lambda(x))), i = 1, 2 \tag{9}$$

where $\mathrm{BN}()$ means batch normalization. Finally, we transpose the feature map $F_{2-\mathrm{i}}$, and multiply them by the feature map G_{3-i} and random initialization coefficient α_1, α_2. Then, the output feature map of spatial attention mechanism module can be obtained:

$$Y = \sum_{i=1}^{2} \alpha_i G_{3-i} F'_{2-i}. \tag{10}$$

Finally, the feature maps obtained by rotating at different angles are averaged to obtain the output feature map of RAM module:

$$Y_{\mathrm{out}} = \frac{1}{K} \sum_{i=1}^{K} Y_i \tag{11}$$

where Y_i is the output feature map obtained after the input feature map is rotated by the angle of λ_i. K is the number of rotation angles.

Because the features extracted from the shallow convolution layer and the middle convolution layer pay more attention to the edges and corners, RAM module is chosen to be inserted after the last ReLU layer of every bottleneck layer of Resnet50 network, which is consistent with the feature that RAM module is used to make the network pay more attention to the semantic information in the ISAR image. The structure of network is as shown in Fig. 7. For ISAR image recognition of space targets, if only a small sample of ISAR data is used to train a new deep CNN model from scratch, it is easy to cause overfitting and other problems. Therefore, referring to the idea of transfer learning (TL), the deep CNN trained on the large-scale Imagenet dataset is transferred to the task of space target ISAR image recognition in order to accelerate the training process, and then use the ISAR image dataset in the target domain to train Resnet50 again until the network matches ISAR image parameters. The size of the input layer image of Resnet50 is 224 $\times$224 pixels. First, the weight of the convolution layer used for feature extraction is frozen, and then the weight of RAM is initialized randomly with Gaussian initialization method, which means only the weights of the fully connected layer and RAM module are trained. Then, a new deep neural network can be obtained by using the ISAR image simulation dataset to train the classification neural network again.

B. Graph Convolutional Network

Due to the weak network generalization ability of a single CNN structure, it makes sense to introduce small-sample classification network such as GCN. For GCN, the embedding of nodes is learned by CNN structure, where the node refers to the image to be recognized. The prediction of node labels in the test set depends on the measurement relationship between its embedding and the embedding of nodes in the training set. Intuitively, GCN intends to find which kind of nodes in test nodes set and training nodes set are more similar, so as to optimize the embedding function. The purpose of graph few-shot learning is to learn an embedding function with good performance from this series of graphs, so that it can be used for new graphs with only a small number of training set images. The different graphs here correspond to different tasks in meta-learning, indicating the generalization of task level. Several important concepts in GCN are as follows:

$G = (V, E, A)$ is a graph data, where V is the set of nodes, E is the set of edges, and A is the adjacent matrix which is used to describe the similarity between the features of adjacent nodes. $v_i \in V$ describes a point. $e_{ij} = (v_i, v_j) \in E$ describes the edge between two nodes. A is an $N \times N$ matrix, where

$$A_{ij} = \begin{cases} w_{ij} & \text{if } \mathrm{e_{ij}} = (\mathrm{v_i}, \mathrm{v_j}) \in \mathrm{E} \\ 0 & \text{if } \mathrm{e_{ij}} \notin \mathrm{E}. \end{cases} \tag{12}$$

Graph is associated with node attribute X, $X \in R^{N\times D}$ is an eigenmatrix, and $X_i \in R^D$ represents the eigenvector of node v_i, where N is the number of nodes and D is the length of the eigenvector. Fig. 8 illustrates the GCN structure of ISAR image classification of space targets in this article. First, train data is input into convolutional neural network for feature extraction. The second part is GCN, which is a GCN, and its goal is to construct a conditional distribution $P(Y|T)$, from which we can infer the classification result Y for a given tensor T. At this time, Y corresponds to the label of the input task data, and forms a supervised learning loss with it. For GCN, the feature extracted from CNN is taken as its input. Each color circle represents each node in the graph, and different colors represent different

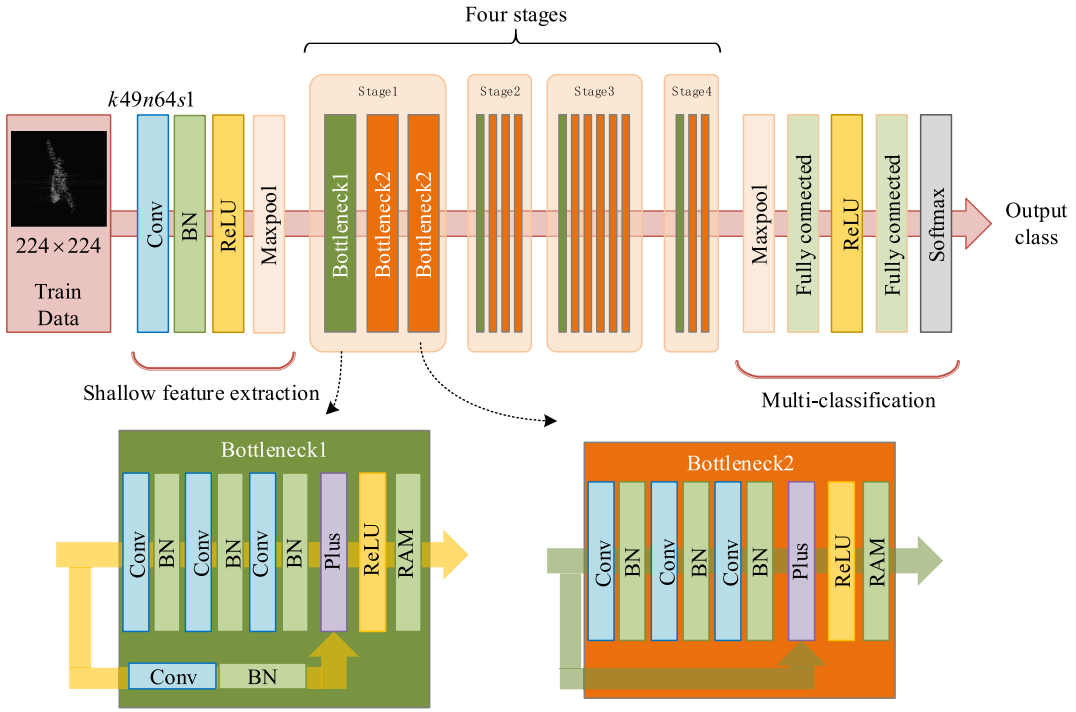


Fig. 7. Network structure after adding RAM module to Resnet50 network.

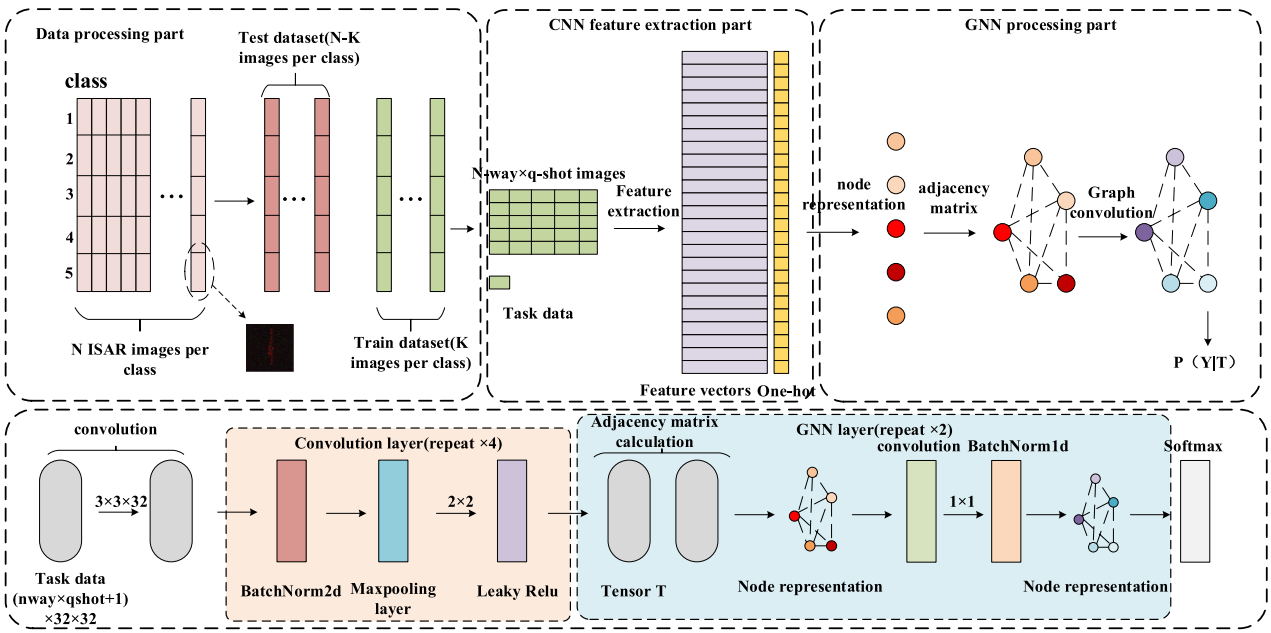


Fig. 8. Graph convolutional network structure.

features of each node. The adjacent matrix is constructed as follows:

$$\tilde{A}_{i,j}^{(k)} = \varphi_{\tilde{\theta}}(x_i^{(k)}, x_j^{(k)}) \tag{13}$$

where $\varphi_{\tilde{\theta}}()$ is multilayer perception which inputs the absolute value difference between two nodes and outputs the corresponding weight value. $x_i^{(k)}$ is the node of layer k, and i represents the length of the eigenvector of the node. The expression of $\varphi_{\tilde{\theta}}()$ is as follows:

$$\varphi_{\tilde{\theta}}(x_i^{(k)}, x_j^{(k)}) = \mathrm{MLP}_{\tilde{\theta}}(\mathrm{abs}(x_i^{(k)} - x_j^{(k)})) \tag{14}$$

where $\mathrm{abs}()$ means a process of taking an absolute value. Each row is processed by Softmax function to ensure that the sum of all weights in this row is 1. Then the processed adjacent matrix

B is obtained. After that, GCN can be used to continue to iterate to the next layer. The calculation process is as follows:

$$x_l^{(k+1)} = G_c(x^{(k)}) = \rho\left(\sum_{B\in \mathrm{A}} Bx^{(k)}\theta_{B,l}^{(k)}\right), l = d_1 \ldots d_{k+1} \tag{15}$$

where ρ is the nonlinear activation layer, in which leaky-ReLU function is used. θ is a learnable parameter, and the multilayer GCN is obtained by repeating the above steps. The cross-entropy loss function is used to calculate the loss:

$$l = -\sum_k y_k \log P(Y_* = y_k \,|\Upsilon) \tag{16}$$

where the model is asked to predict the label Y corresponding to the image to classify $x \in \Upsilon$. $*$ represents a node in the graph. In graph convolution operation, the process of finding depth feature relationship can be understood as follows: In the graph, the balls of various colors represent the deep features of convolution mining. The relationship between deep features is mined by graph convolution to produce different node representations, so the color of the ball changes. In Fig. 8, the blue ball retains more feature information than the red ball and will be retained as an important feature for model prediction. In the last step, the purple ball and the blue ball are more closely related, so they are finally retained. The line between the two balls represents the relationship between them, which provides more information for the final classification.

C. Relation Network

The disadvantage of GCN is its poor flexibility and scalability, and when the number of space target types is small the classification effect will not be good. In order to overcome the above defects and enhance the generalization ability of the network, it is helpful to introduce RN. In the meta-learning process, network learns the depth distance metric to compare a small number of images in the series, and each episode is designed to simulate the setting of few shots. Once trained, the network can classify the images of the new category by calculating the relationship score between the query image and several examples of each new category without further updating the network. In the field of small sample learning, RN can be considered as a method based on metric. However, different from other measurement-based methods, RN uses neural network to learn this measurement. The embedded module generates the representation of query and training image, which is depth feature. Then, the relational module compares these embeddings to determine whether they come from the matching category.

The network structure of ISAR image recognition based on RN is shown in Fig. 9. The sample x_i in the training set and the sample x_j in the validation set are input into the module f_φ to generate the feature maps $f_\varphi(x_i)$ and $f_\varphi(x_j)$. They are combined by function $C()$ in depth. Then, the combined feature map is fed to the relation module g_ϕ, which finally generates a scalar in the range of 0–1. And this scalar is called relation score, which represents the similarity between x_i and x_j. The

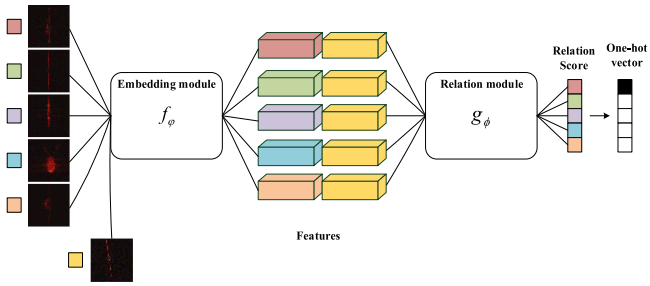


Fig. 9. Relation network structure.

expression of relation score is as follows:

$$r_{i,j} = g_\phi(C(f_\varphi(x_i), f_\varphi(x_j))), i = 1, 2, \ldots..K \tag{17}$$

where K is the number of types of space targets randomly selected. Finally, a one-hot encoding is generated from the relationship score, which represents the final classification result of the network. The loss function adopts the MSE loss function, which can be expressed as

$$L = \arg\min_{\phi,\varphi} \sum_{i=1}^{m}\sum_{j=1}^{n}(r_{i,j} - 1)^2 \tag{18}$$

where 1 means the final relationship score of matching pairs and we give a 0 for those mismatched image pairs. Different from the general measurement classification method, the RN calculates the measurement through the network learning method, and sets the loss function to continuously optimize the feature embedding process, which is the reason why RN can realize small sample classification.

D. Meta-Learner Based on Adaptive Weighted XGBoost

Different base learners have different biases in image feature selection and recognition. The combination of Resnet50 and RAM module pays more attention to the semantic information of ISAR images of space target. The GCN searches the deep feature relationship of ISAR fine-grained image of space target by constructing the neighborhood matrix between nodes, and the RN searches the same category of images by calculating the distance between image features. Therefore, the meta-learner should be chosen as a model with strong generalization ability to correct the bias of multiple base learners on the training data. When the data is discrete, XGBoost can be used to solve multiclassification problems. However, for the XGBoost algorithm, the training data is generally based on the same weight. For the case of multiple learners in this article, the weights of the meta-features obtained should be different and adaptive. Therefore, this article proposes an adaptive weighted XGBoost algorithm to solve the multilearner integration problem.

$D\{\mathrm{x}_{i,\alpha}, l_i, \mathrm{e}_{i,\alpha}\}(i = 1, 2, \ldots, \mathrm{n}, \alpha = 1, 2, 3)$ extracted by base learners is used as input for the tree model, where $x_{i,\alpha}$ is the meta-feature of the *i*th sample in the new training set, l_i is the label corresponding to the *i*th sample in the new training set which is represented by one-hot encoding, $e_{i,\alpha}$ is the weight corresponding to the meta-feature, and α is the order of the base

learners. The expression of $e_{i,\alpha}$ is as follows:

$$e_{i,\alpha} = \frac{(x_{i,\alpha,1}-\overline{x_{i,\alpha}})^2+(x_{i,\alpha,2}-\overline{x_{i,\alpha}})^2+\ldots+(x_{i,\alpha,n}-\overline{x_{i,\alpha}})^2}{n} \cdot q_\alpha \quad (19)$$

where $x_{i,\alpha,n}$ is the *n*th data in the meta-feature, representing the confidence of the αth base learner for the classification result of the *n*th category. *n* is equal to the number of categories of the space targets. $\overline{x_{i,\alpha}}$ is the mean value of the data in the meta feature. q_α is the weight corresponding to the base learners. $\hat{y}_{i,\alpha}$ is the predicted value of the *i*th sample, whose dimension is the same as the number of categories of space target:

$$\widehat{y}_{i,\alpha} = e_{i,\alpha} \bullet \phi(x_{\ i}) = \sum_{k=1}^{K} e_{i,\alpha} \bullet f_k(x_i), f_k \in F \quad (20)$$

where $F = \{f(x) = w_{q(x)}\}$ is the space of classification tree, and q represents the structure of each tree, w represents the weight of leaves of each tree; K is the number of decision trees used; f_k represents the *k*th tree model, and each f_k corresponds to an independent tree structure and the weight of leaf nodes. The predicted value of $\hat{y}_{i,\alpha}$ is the final classification result of the meta-learner. The optimization process of XGBoost is to minimize the following loss functions:

$$L = \sum_i l(\hat{y}_{i,\alpha}, l_i) + \sum_k \Omega(f_k) \quad (21)$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$$

l is a differentiable convex loss function used to measure the difference between the predicted value $\hat{y}_{i,\alpha}$ and the label l_i. $\Omega()$ is a regularizer which can help to smooth the final learnt weights to avoid overfitting. T is the number of leaf nodes of the tree. γ is the regularization parameter of the number of leaves, and λ is the regularization parameter of the weights of leaves. Softmax cross-entropy loss function is used for multicategory classification here, which can be represented as

$$l(l_i, \widehat{y}_{i,\alpha}) = -l_i \bullet \log\left(\frac{e^{\hat{y}_{i,\alpha}}}{\sum_j e^{\hat{y}_{j,\alpha}}}\right). \quad (22)$$

The XGBoost algorithm needs to use the first derivative and the second derivative of the loss function, and their expressions are derived as

$$l'(l_i, \hat{y}_{i,\alpha}) = y_{i,pred,\alpha} - l_i \quad (23)$$

where

$$y_{i,\text{pred},\alpha} = \frac{e^{\hat{y}_{i,\alpha}}}{\sum_j e^{\hat{y}_{j,\alpha}}}$$

and

$$l''(l_i, \widehat{y}_{i,\alpha}) = y_{i,\text{pred},\alpha} \bullet (1 - y_{i,\text{pred},\alpha}). \quad (24)$$

For $y_{i,\text{pred},\alpha}$ in the first tree, the conventional XGBoost algorithm sets it to 0.5, but for the adaptive weighted XGBoost algorithm in this article, since the value in the meta-feature has a direct physical meaning, the initial value is set to the largest predicted probability value, which can be represented as

$$y_{i,\text{pred},\alpha,\text{initial}} = \text{argmax}(\hat{y}_{1,\alpha}). \quad (25)$$

The advantage of this setting is that it can make the tree split faster. In the process of optimization, the tree model should minimize the loss function as much as possible. The greedy algorithm is used to iteratively add branches to the tree, starting from a leaf. The specific optimization process of XGBoost can be seen in [49]. In order to obtain the optimal tree structure, the maximum gain is calculated in each iteration. In the adaptive weighted XGBoost algorithm in this article, the method of gain calculation is as follows:

$$\text{Gain} = 0.5\left[\frac{G_L^2}{H_L^2+\lambda} + \frac{G_R^2}{H_R^2+\lambda} - \frac{(G_L+G_R)^2}{(H_L+H_R)^2+\lambda}\right] - \gamma \quad (26)$$

where

$$G_L = \sum_{\alpha=1}^{3}\sum_{i\in I_L} e_{i,\alpha} \bullet g_{i,\alpha}, H_L = \sum_{\alpha=1}^{3}\sum_{i\in I_L} e_{i,\alpha} \bullet h_{i,\alpha}$$

$$G_R = \sum_{\alpha=1}^{3}\sum_{i\in I_R} e_{i,\alpha} \bullet g_{i,\alpha}, H_R = \sum_{\alpha=1}^{3}\sum_{i\in I_R} e_{i,\alpha} \bullet h_{i,\alpha}$$

where $g_{i,\alpha}$ and $h_{i,\alpha}$ represent the first derivative value and the second derivative value of the *i*th meta feature obtained by αth base learner. I_L and I_R represent the samples contained in the left node and the samples contained in the right node, respectively. In this way, the meta-features obtained by the three base learners can be used to adjust the weight of the meta-learner.

The size of q_α is adaptively changed with training. The advantage is that it can reflect the contribution of the base learners to recognition more concretely, and it can also enhance the final recognition effect. The initial values of q_1, q_2, and q_3 are all set to 1/3, whose purpose is to make the base learners contribute the same to the recognition result at the beginning of the training process. First, the partial derivative of the loss function L with respect to q_α is calculated, and the result is as follows:

$$\frac{\partial L}{\partial q_\alpha} = \frac{\partial L}{\partial l} \bullet \frac{\partial l}{\partial \hat{y}_{i,\alpha}} \bullet \frac{\partial \hat{y}_{i,\alpha}}{\partial e_{i,\alpha}} \bullet \frac{\partial e_{i,\alpha}}{\partial q_\alpha} = \left(\frac{e^{\hat{y}_{i,\alpha}}}{\sum_j e^{\hat{y}_{j,\alpha}}} - l_i\right) \bullet \phi(x_i) \bullet p_{i,\alpha} \quad (27)$$

where

$$p_{i,\alpha} = \frac{(x_{i,\alpha,1}-\overline{x_{i,\alpha}})^2 + (x_{i,\alpha,2}-\overline{x_{i,\alpha}})^2 + \ldots + (x_{i,\alpha,n}-\overline{x_{i,\alpha}})^2}{n}.$$

Then, use the following formula to update the value of q_α:

$$q_\alpha^{(k+1)} = q_\alpha^{(k)} - \eta \bullet \frac{\partial L}{\partial q_\alpha} \quad (28)$$

where $q_\alpha^{(k)}$ represents the value of q_α at the *k*th iteration. η is the update rate, and is set to 0.1 here.

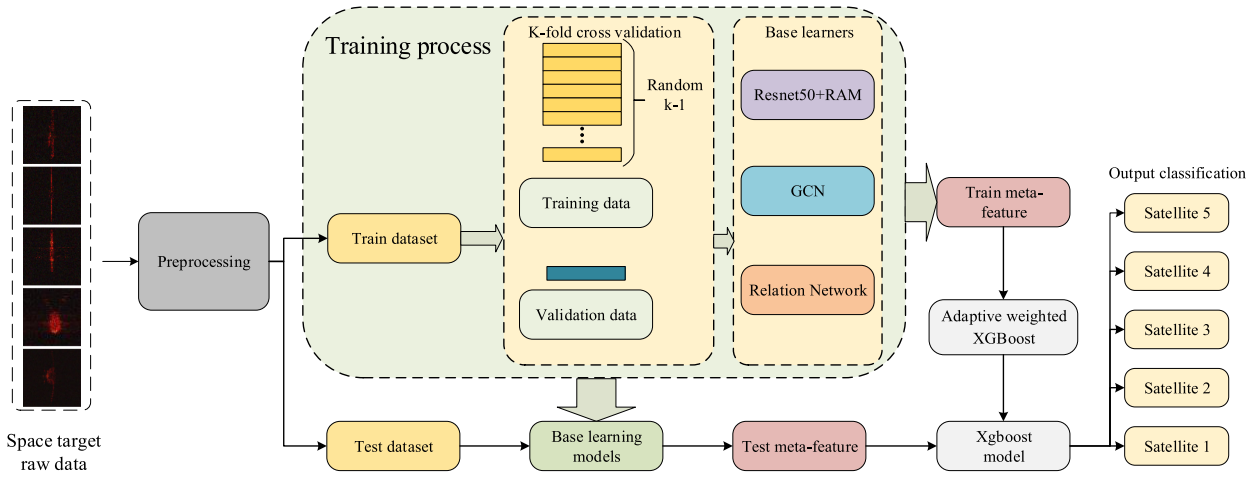


Fig. 10. Overall process of the classification method in this article.

The overall flow of classification method in this article is shown in Fig. 10, and the detailed steps of the method are as follows:

First, the method preprocesses the original ISAR dataset, including filtering, power transformation, energy normalization, and data augmentation (DA). Then, the whole dataset is divided into train dataset $S_{\text{train}} = \{(x_{\text{tr}}, l_{\text{tr}}), \text{tr} = 1, 2, \ldots . N_{\text{tr}}\}$ and test dataset $S_{\text{test}} = \{(x_{\text{te}}, l_{\text{te}}), \text{te} = 1, 2, \ldots . N_{\text{te}}\}$, where N_{tr} is the number of train dataset and N_{te} is the number of test dataset. Because the total number of samples is small, the idea of cross-validation is used to randomly divide the train dataset into k equal blocks $S_{\text{tr1}}, S_{\text{tr2}}, \ldots . S_{\text{tr}K}$, which have no cross parts. K is determined by the total number of data, which is corresponding to the total number of the whole dataset, and here 10 is chosen.

Next, select one block S_{tri} as the validation dataset, and the rest $k-1$ blocks as the training dataset, and then send them to three base learners for training. Repeat the above steps for K times to get the trained base learner L_{trained}. The meta-features of the training dataset are obtained by combining the classification results of validation dataset, and the meta-features of the test set are obtained by combining the classification results of the test dataset.

Finally, the meta-features of the training dataset and the real classification results are input to train the meta-learner to get the trained meta-learner M_{trained}. The meta-features of the test dataset are input into M_{trained} to get the final more accurate classification results. Algorithm 1 gives the detailed steps.

Algorithm 1: MSN For ISAR Image Recognition.

Input:
ISAR image dataset S; depth of the tree T.
Output:
Predict label for input image.
Algorithm:
1: Preprocess input dataset.
2: Divide S into train dataset S_{train} and test dataset S_{test}.
3: Divide S_{train} into K equal blocks, $S_{\text{tr1}}, S_{\text{tr2}}, \ldots . S_{\text{trK}}$.
4: Use *K*-fold cross-validation method to train the base learners and get the training models.
5: Combine the classification probabilities of S_{train} to get train meta-feature f_{train}.
6: Combine the classification probabilities of S_{test} to get test meta-feature f_{test}.
7: Train adaptive weighted XGBoost by f_{train}.
8: **for** $t = 1$ to T do
9: Calculate loss function L and gain function Gain according to (21) and (26).
10: Calculate the partial derivative of L with respect to q_α according to (27).
11: Update the value of q_α and the weight $e_{i,\alpha}$.
12: **end for**
13: Calculate the network output by f_{test} and conduct label prediction.

IV. Experiment

First, a simulation dataset will be used for classification experiments. The method proposed in this article will be compared with the current mature methods by calculating the kappa parameter of the obtained confusion matrix and analyzing the impact of the RAM module on the Resnet50 network through the visualization method of Grad-CAM intuitively [51]. Then, a histogram of the influence weight of the features extracted by adaptive weighted XGBoost algorithm is given and the reason why the classification effect is better than that of a single base learner is analyzed. In the end, we add the measured ISAR images of Tiangong-1 published by German FGAN Lab in 2018 (@ Fraunhofer FHR) into the simulation dataset and perform experiments. Through the analysis of the classification confidence, the superiority of the MSN method compared with similar methods is proved.

A. Experiment Dataset

When building the simulated ISAR image dataset of space target, SolidWorks software is first used to build the geometric

TABLE I
DATA VOLUME IN DIFFERENT *K/N*

K/N \ Category	Original Dataset			Dataset after DA		
	Total Num	Train Num	Test Num	Total Num	Train Num	Test Num
0.2	200	33	167	800	132	668
0.7	200	82	118	800	328	472
1.2	200	109	91	800	436	364

model of space target: satellite 1, satellite 2, satellite 3, satellite 4, and satellite 5. Then, FEKO software uses the material of target, wavelength of electromagnetic wave [52], shape of target, path of current flowing through target, and other parameters to analyze the shape of magnetic field and electric field around antenna and other target objects. The intensity and other electromagnetic parameters are simulated, and the satellite model is meshed, with the frequency of 1 GHz. The corners of the grid are extracted as the scattering points of the target.

The parameters of the RD algorithm used in the simulation image in this article are as follows: in order not to make the resolution too high, the scattering points are set to 500, and the initial angle between the reference coordinate system and the local coordinate system is $-130°$. The space target is illuminated by pulse linear frequency modulation signal with frequency $f_c =$ 5.52 GHz, bandwidth $B = 300$ MHz, pulsewidth $T_p = 25.6$ μs, and range sampling frequency $F_M = 2.5$ MHz. The target rotation angular velocity is 0.1 $\mathrm{rad/s}$. The target is imaged in different rotation periods, and 200 target ISAR images are obtained to simulate the observation results at different angles, as shown in Fig. 11. This article simulates the recognition results under different ratios of training data number to test data number. The ratio is expressed by *K/N*, where *K* is the number of training samples and *N* is the number of test samples. Table I shows the number of original dataset and dataset after DA under different *K/N* conditions.

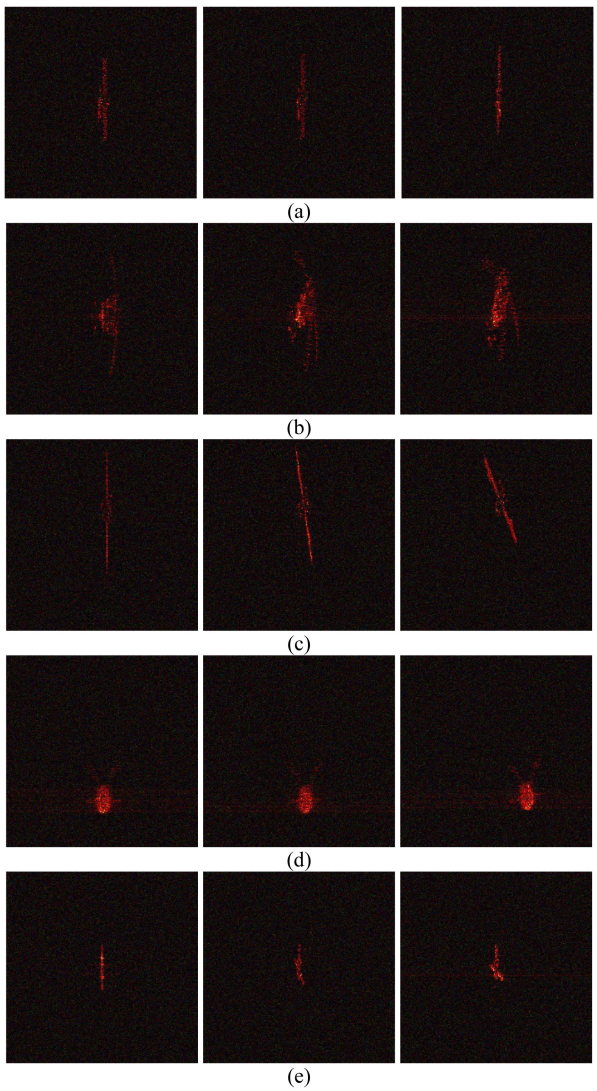


Fig. 11. ISAR image simulation results of space target. (a) sat1. (b) sat2. (c) sat3. (d) sat4. (e) sat5.

B. Evaluation Protocol

The experiment uses the classification accuracy (ACC), cross-entropy loss, confusion matrix, and kappa coefficient to evaluate the classification performance of the model. The calculation method is as follows:

$$\begin{aligned} \mathrm{ACC} &= N_{\mathrm{corr}}/N_{\mathrm{total}} \\ \mathrm{Loss} &= -\sum_{i=1}^{N_{total}}\sum_{j=1}^{K} l_{ij}\log l_{ij} \\ \mathrm{kappa} &= \frac{OA - p_e}{1 - p_e} \end{aligned} \tag{29}$$

where N_{corr} and N_{total} are the number of images correctly recognized and total number, respectively. l_{ij} is prediction probability, and i and j are the image serial number and the learner serial number, respectively. p_e is the relative misclassification number. Assuming that the number of real samples of each class is $A_1, A_2, \ldots A_c$, respectively, and the number of predicted samples of each class is $B_1, B_2, \ldots B_{\mathrm{c}}$, respectively, and the total number of samples is *N*, then

$$p_e = \frac{A_1 \times B_1 + A_2 \times B_2 + \ldots + A_c \times B_c}{N \times N}. \tag{30}$$

The kappa coefficient is a measure of classification accuracy. Its calculation is based on a confusion matrix. The value range is between 0 and 1. The larger the value, the higher the classification accuracy. For the three base learners and the MSN method, the confusion matrix of the classification results of five space targets is calculated. To ensure the consistency of the experiment, *K/N* is set to 1.2. For the test dataset, the larger ACC and the

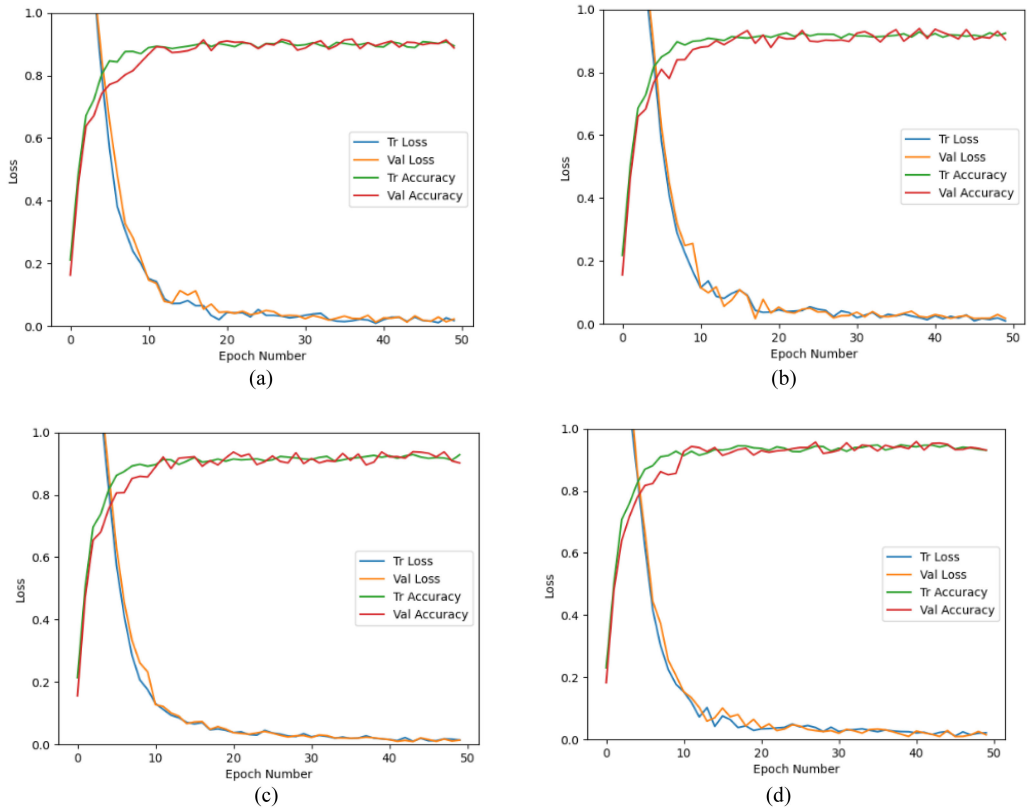


Fig. 12. Loss and accuracy curve. (a) Resnet50. (b) Resnet50 + RAM. (c) RN. (d) GCN.

smaller cross-entropy loss mean the stronger data classification ability of the model.

C. Network Details

Pytorch is used to implement the proposed framework. All the experiments were carried out on Intel Core i7 with 3.39 GHz CPU and NVIDIA Quadro p4000 GPU.

1) Resnet50 + RAM Module: The Resnet50 model pre-trained on Imagenet dataset is used for TL, and then finetuned on the small sample space target dataset. The article uses the random gradient descent [53] optimizer and sets the momentum of the model to 0.9 and the weight decay to $4e^{-5}$. The learning rate before the fully connected layer is set to 0.0001, and the learning rate of the fully connected layer is set to 0.001, so that the change of feature extraction is slightly slower, and the result converges more fully. According to Table I, when $K/N = 1.2$, the number of test dataset is 91 for each type. The number of epoches is set to 50, and batchsize is set to 32. Training accuracy, validation accuracy, cross-entropy loss during training, and cross-entropy during validation are shown in Fig. 11. When each type of training images contains 109 pieces, each round of training takes about 21 s, and the total training time is 17.5 min. It can be seen from Fig. 12 that, as the number of training epoches increase, the training accuracy and validation accuracy increase, while the training cross-entropy loss and validation cross-entropy loss decrease. At the end of the training, the training accuracy and validation accuracy are $94.46 \pm 0.22\%$ and $92.44 \pm 0.44\%$, respectively, and the training cross-entropy loss and validation cross-entropy loss are 0.0051 and 0.0078, respectively.

2) Graph Convolutional Network: The GCN structure as shown in the figure above is adopted. The five types of data samples are input into the network according to the ratio of training dataset to test dataset. During training, N pictures are selected for four types of all satellite data, $N+1$ pictures are selected for the remaining fifth type, and the extra one is used as validation. The features extracted by the CNN are connected with a one-hot code as the input of the GCN. The batchsize is set to 16, the learning rate of the whole process is set to 0.01, and the loss function adopts the cross-entropy loss function. In order to adapt to the network input, each image is compressed into a size of 32×32 pixels in the network. Under the condition of $K/N = 1.2$, after 12 000 iterations, the training accuracy and validation accuracy are $98.90 \pm 0.22\%$ and $94.46 \pm 0.22\%$, respectively, and the training cross-entropy loss and validation cross-entropy loss are 0.0079 and 0.0102, respectively.

3) Relation Network: The five types of satellite data are sent to the RN at the same set ratio. The size of each image in the network is resized to 84×84 pixels. One point worth noting in the RN is that in each test, at least 20 images need to be sent into the network which can guarantee the accuracy of the test because the relationship network relies on calculating the distance between similar sample data to realize small sample recognition. If there are too few data, incomplete distance calculation may lead to poor recognition results. The batch size is set to 10, the number of training iterations is 50 000, the number of test

TABLE II
DETAILED NET PARAMETERS

Model	Resnet50 +RAM	GCN	RN	XGBoost

Layers/ Parameters	Layer0{
	Cov(64)
	BN(64)
	ReLU
	MP(2×2)}
	Layer1{
	3×Bottleneck{
	Cov(64)
	BN(64)
	Cov(64)
	BN(64)
	Cov(256)
	BN(256)
	ReLU
	RAM}}
	Layer2{
	4×Bottleneck}
	Layer3{
	6×Bottleneck}
	Layer4{
	3×Bottleneck}
	Out{
	AvgPool (1)
	Linear(256)
	ReLU
	Linear(5)
	Softmax(5)}
		CNN embedding{
		Cov(32)
		BN(32)
		MP(2×2)
		ReLU
		Cov(64)
		BN(64)
		MP(2×2)
		ReLU
		BN(128)
		MP(2×2)
		ReLU}
		GCN{
		Adjacency_layer{
		Cov(64)
		BN(64)
		ReLU
		Cov(32)
		BN(32)
		ReLU
		Cov(1)}
		Graph_conv_block{
		Linear(16)
		BN(16)
		Linear(5)}}
			CNNEncoder{
			Cov (64)
			BN(64)
			ReLU
			MP(2×2)
			Cov (64)
			BN(64)
			ReLU
			Cov (64)
			BN(64)
			ReLU
			Cov (64)
			BN(64)
			ReLU}
			RelationNetwork{
			Cov (64)
			BN(64)
			ReLU
			MP(2×2)
			Cov(64)
			BN(64)
			ReLU
			MP(2×2)
			Linear(8)
			Linear(5)}
				max_depth: 20,
				num_boost_round:2000
				objective: multi:softprob
				random_state: 7,
				silent:0,
				num_class:5,
				ESR:0.8

Lr	0.0001（base） 0.001（fc）	0.01	0.001	0.4
batchsize	32	16	10	-
Input size	224×224	32×32	84×84	1×5
Iteration	2800	12000	50000	2000

Max pooling (MP), the BatchNorm module (BN), the ReLU function (ReLU), the rotation-invariant attention mechanism module (RAM).

iterations is 600, the learning rate of the whole process is set to 0.001, and the hidden layer unit size is set to 10. At the end of training, the training accuracy and validation accuracy are, respectively, 96.44±0.22% and 94.68±0.44%, and the training cross-entropy loss and validation cross-entropy loss are 0.0069 and 0.0098, respectively.

4) Adaptive Weighted XGBoost: The input of XGBoost is the classification probabilities of the five types of satellite targets, the true labels and the weights corresponding to the meta-features. The learning rate is set to 0.4, and the depth of the tree is chosen to be an appropriate value. Twenty is chosen as the depth of the tree because too large depth may cause overfitting. The number of iterations is set to 2000, and the target type is set to "multi:softprob," which means that the target problem is a multiclass problem. The "num-class" parameter is set to 5, which is consistent with the number of categories of the training data. In order to prevent overfitting, it makes sense to set the "early_stopping_rounds" (ESR) parameter to be used in the update process. ESR reduces the weight of the feature and makes the calculation process more conservative, which is set to 0.8. When *K*/*N* is equal to 1.2, after training with adaptive weighted XGBoost algorithm, the final test accuracy is 99.78±0.22%, which is about 5% higher than the average classification accuracy of the three base learners.

Table II lists the specific network structure of the three base learners, the settings of adaptive weighted XGBoost, batchsize, learning rate, and other parameters. Among them, Cov(64) means that the number of output layers of the convolutional layer is 64. For Resnet50 network with RAM module, the RAM module is embedded in the bottleneck layer, which is clearly shown in the table. The last three base classifiers all output five values, corresponding to the classification probabilities of the five types of satellite data.

D. Test Results and Analysis

The test accuracy listed in Table III is under the conditions of $K/N = 0.2$, 0.7, and 1.2. In order to prove the effectiveness of MSN method, Alexnet, MAML [54], siamese network, matching network, and prototypical network are used to compare the recognition results under the same amount of data and test conditions. With the increase of *K*/*N*, the recognition accuracy of various methods has been improved. When *K*/*N* is equal to 1.2, the classification accuracy of MSN reaches the maximum, which is 99.78±0.22%. It has a competitive advantage in similar classification methods, which proves that this method is indeed effective. The results show that in the case of different amount of training data, the accuracy of MSN

TABLE III
TEST ACCURACY OF EACH CATEGORY OF DATA CLASSIFICATION UNDER DIFFERENT METHODS (%)

OA/(K/N)	0.2	0.7	1.2
Fine-tuned Alexnet	70.24±0.44	80.90±0.44	89.56±0.44
MAML	72.22±0.22	80.24±0.22	90.24±0.22
Siamese Nets	74.68±0.44	83.58±0.22	92.44±0.22
Matching Nets	75.56±0.22	84.68±0.44	94.68±0.44
Prototypical Nets	78.90±0.22	88.68±0.22	98.90±0.22
MSN(ours)	**82.44±0.22**	**91.34±0.44**	**99.78±0.22**

The bold values means our method MSN, and in order to have a contrast with other methods, we marked our method in bold values.

TABLE IV
CONFUSION MATRIX OF RESNET50 + RAM

True/Predict	1	2	3	4	5
1	0.83	0.06			
2		0.94		0.06	
3			0.96		
4	0.15		0.02	0.94	0.07
5	0.02		0.02		0.93

TABLE V
CONFUSION MATRIX OF GCN

True/Predict	1	2	3	4	5
1	0.91			0.04	
2		0.96			
3			1.0		
4	0.06	0.04		0.96	0.03
5	0.03				0.97

TABLE VI
CONFUSION MATRIX OF RN

True/Predict	1	2	3	4	5
1	0.89			0.06	
2		0.92			0.03
3	0.09		0.98		
4		0.08		0.94	
5	0.02		0.02		0.97

TABLE VII
CONFUSION MATRIX OF MSN

True/Predict	1	2	3	4	5
1	0.97			0.01	
2		0.99			
3			1.00		
4	0.03			0.99	
5		0.01			1.00

method is about 5% higher than that of GCN and other base learners. Tables IV–VII, respectively, list the confusion matrix of Resnet50+RAM module, GCN, RN, and MSN. According to formula (29), their corresponding kappa coefficients can be calculated, which are 0.92, 0.96, 0.94, and 0.99, respectively, and the kappa coefficients can prove the consistency of the classification results of the MSN method is better and the overall accuracy is higher.

It is also shown from Table III that adding RAM module to fine-tuned Resnet50 network can improve the accuracy. Grad-CAM method is used to multiply the feature map extracted from the last layer of convolution layer by the final classification probability, and express it in the form of heat map, where the red part represents the focused part of the network and the blue part represents the ignored part of the network. Fig. 13(a) shows the original image of the five types of ISAR images, Fig. 13(b) shows the heat map extracted by the Grad-CAM method after RAM module is added to the Resnet50 network, and Fig. 13(c) shows the extracted heat map by Resnet50 network. As shown in the figure, through the effect of RAM module, the network focuses on the space target part in the ISAR image, while weakening the influence of the irrelevant black background. Through this module, for different kinds of ISAR images, the parts of attention become more detailed, rather than distracted.

E. Adaptive Weighted XGBoost Result Analysis

The above results show that after the training of adaptive weighted XGBoost, the classification accuracy has indeed improved. In order to illustrate the effectiveness of adaptive weighted XGBoost more intuitively, this article scores the importance of five meta-features and compares with the results of the general XGBoost algorithm. Fig. 14(a) shows the training feature scores of adaptive weighted XGBoost algorithm. f0 to f4 are the predicted probabilities of the base learner. It can be seen from the figure that the feature score of f0 is the lowest, which is 28. The feature score of f3 is slightly higher than that of f0, which is 35. The scores of f2, f1, and f4 are the highest, 50, 54, and 58, respectively, indicating that through the training of adaptive weighted XGBoost, the first and fourth types of features contribute less to the final classification results. Because the adaptive weighted XGBoost algorithm analyzes that it is just these two types of features that lead to a low classification rate, as shown in the confusion matrix from Tables IV–VII. It can be seen that these two categories indeed reduce the classification probabilities of the base learners. Fig. 14(b) shows the training feature scores of general XGBoost algorithm. It can be seen that the feature importance by general XGBoost algorithm is more concentrated than that of adaptive weighted XGBoost algorithm. The variances of the scores of the two algorithms are 132.8 and 36.8, respectively, which illustrates that adaptive weighted XGBoost algorithm distinguishes the features of the target better and has better classification performance. Finally, after the training of adaptive weighted XGBoost, the classification results of the three base learners are combined to improve the recognition rate and prove the effectiveness of the method.

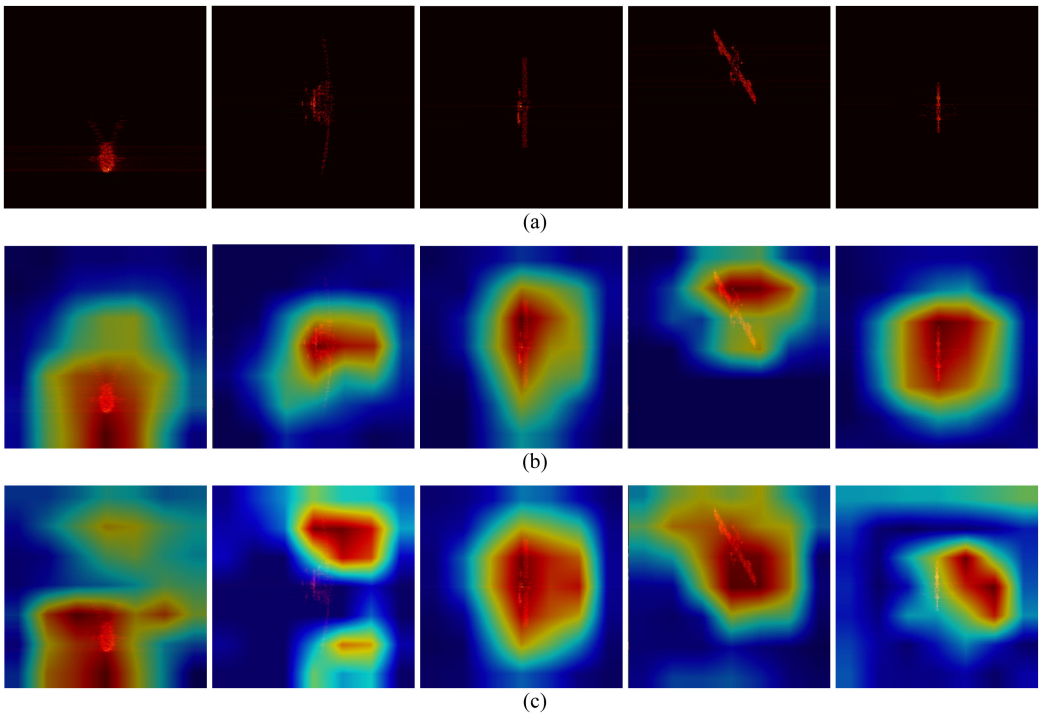


Fig. 13. Attention heat map obtained by Grad-CAM method. (a) ISAR original image of satellite 1 to satellite 5. (b) Attention heat map after adding RAM module. (c)Attention heat map without RAM module.

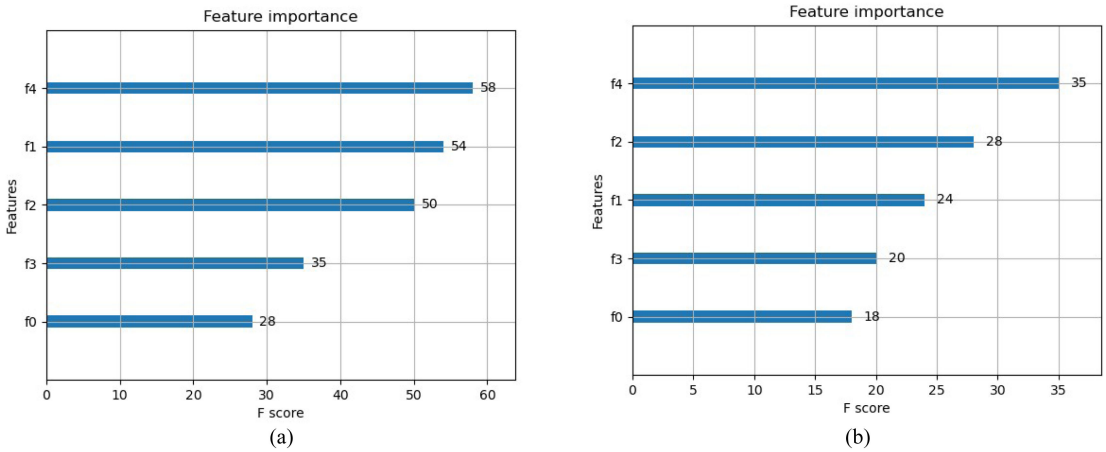


Fig. 14. Feature importance distribution. (a) Adaptive weighted XGBoost algorithm. (b) General XGBoost algorithm.

F. Measured Data Experiment

In this experiment part, the measured data is used to conduct experiments to highlight the contributions of the method in this article. The ISAR image dataset is the result of long-time observations of Tiangong-1 with the TIRA system, which was published by German FGAN Lab in 2018 (@ Fraunhofer FHR) [55]. Several measured ISAR images in different postures are shown in Fig. 15. In the experiment process, because there is only one type of measured data of the target, we replace the first type of simulated image with 200 measured ISAR images with different poses, then use the process in Algorithm 1 to classify the new dataset, and compare the classification effect of the similar methods in Table III on the measured data. After many

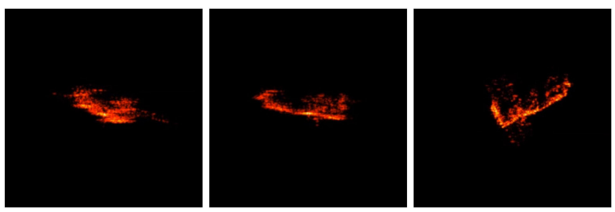


Fig. 15. Measured ISAR images in different postures.

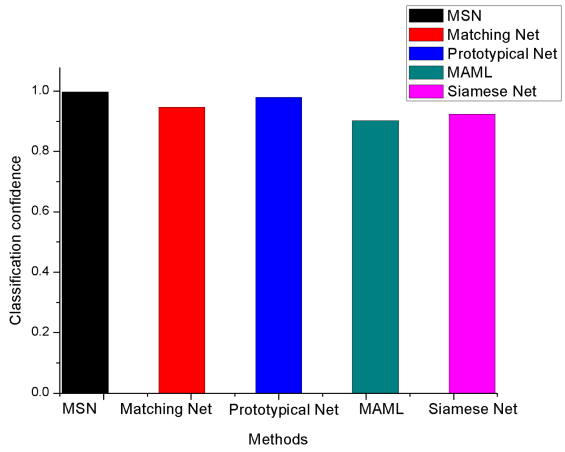


Fig. 16. Classification confidence of different methods.

experiments, the classification confidence obtained is shown in Fig. 16. It can be seen that these methods all have high confidence in the classification of the correct category, exceeding 90%. Among them, the MSN method in this article has the highest correct classification rate, indicating that the method in this article indeed effectively improve the ISAR image recognition under the condition of small samples, and has superiority in similar methods.

V. Conclusion

This article proposes a new method for small-sample recognition of space target by ISAR. The innovative adaptive weighted XGBoost meta-learner improves the recognition accuracy of the base classifiers, and quantitative analysis proves its effectiveness. The problem that deep learning method requires a lot of effective data is solved by adding RAM module into the layers of Resnet50 network. Through the Grad-CAM visualization method, it is proved that the network's attention is indeed more focused on the target. This article has conducted experiments on both the simulation dataset and the measured data. In the experiment on the simulation dataset, under the condition of *K/N* equal to 1.2, the classification accuracy of the MSN method reaches 99.78±0.22%, which is more competitive than similar classification methods. In the experiment on the measured data, it can also be seen that the MSN method in this article has a higher confidence in the correct classification of the measured data than other similar methods, which proves that the method in this article indeed effectively improve the ISAR image recognition effect under the condition of small samples, and has superiority in similar methods. In future research, how to adjust the network structure to adapt to the characteristics of ISAR images will be focused.

References

[1] G. Xu, M. Xing, L. Zhang, Y. Lui, and Y. Li, "Bayesian inverse synthetic aperture radar imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1150–1154, Nov. 2011.

[2] Y. Li *et al.*, "Inverse synthetic aperture radar imaging of ship target with complex motion," *IET Radar, Sonar Navigation*, vol. 2, no. 6, pp. 395–403, 2008.

[3] V. C. Chen and M. Martorella, *Inverse Synthetic Aperture Radar*, vol. 55. Raleigh, NC, USA: SciTech Publishing, 2014, p. 56.

[4] M. Wielgo, P. Samczyński, M. Malanowski, K. Ndini, K. Kulpa, and P. Baranowski, "The SARENKA SAR system-experimental results of ISAR imaging," in *Proc. IEEE 15th Int. Radar Symp.*, Jun. 2014, pp. 1–4.

[5] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 215–223.

[6] J. A. Golden, "Deep learning algorithms for detection of lymph node metastases from breast cancer: Helping artificial intelligence be seen," *JAMA*, vol. 318, no. 22, pp. 2184–2186, 2017.

[7] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[8] S. Chen and H. Wang, "SAR target recognition based on deep learning," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal.*, Nov. 2014, pp. 541–547.

[9] Y. H. Liu, "Feature extraction and image recognition with convolutional neural networks," *J. Phys., Conf. Ser.*, vol. 1087, no. 6, 2018, Art. no. 062032.

[10] W. Ning, W. Chen, and X. Zhang, "Automatic target recognition of ISAR object images based on neural network," in *Proc. Int. Conf. Neural Netw. Signal Process.*, 2003, pp. 373–376.

[11] S. Wagner, "Combination of convolutional feature extraction and support vector machines for radar ATR," in *Proc. 17th Int. Conf. Inf. Fusion*, Jul. 2014, pp. 1–6.

[12] J. Oechssler and A. Roomets, "Unintended hedging in ambiguity experiments," *Econ. Lett.*, vol. 122, no. 2, pp. 243–246, 2014.

[13] W. Hua *et al.*, "Using two formal strategies to eliminate ambiguity in poetry text," in *Proc. Int. Conf. Intell. Sci.*, 2018, pp. 159–166.

[14] W. Huangfu *et al.*, "Radar: A 3D-reram based DNA alignment accelerator architecture," in *Proc. 55th Annu. Des. Automat. Conf.*, 2018, pp. 1–6.

[15] S. Woo *et al.*, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[16] H. Yang, Y. Zhang, and W. Ding, "Multiple heterogeneous P-DCNNs ensemble with stacking algorithm: A novel recognition method of space target ISAR images under the condition of small sample set," *IEEE Access*, vol. 8, pp. 75543–75570, Apr. 2020, doi: 10.1109/ACCESS.2020.2989162.

[17] J. Feng *et al.*, "Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1149.

[18] P. Melgarejo *et al.*, "Leveraging directional antenna capabilities for fine-grained gesture recognition," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 541–551.

[19] M. Martineau *et al.*, "A convolutional neural network into graph space," 2020, *arXiv:2002.09285*.

[20] Q. He, H. Cai, Z.-Z. Han, and C.-X. Shang, "ISAR target recognition based on non-linear manifoldlearning," *Acta Electronica Sinica*, vol. 38, no. 3, pp. 585–590, 2010.

[21] H. Yang, Y. Zhang, and W. Ding, "A fast recognition method for space targets in ISAR images based on local and global structural fusion features with lower dimensions," *Int. J. Aerosp. Eng.*, vol. 2020, 2020, Art. no. 3412582.

[22] A. Elyounsi, H. Tlijani, and M. S. Bouhlel, "Shape detection by-mathematical morphology techniques for radar target clas-sification," in *Proc. 17th Int. Conf. Sci. Techn. Autom. Control Comput. Eng.*, 2016, pp. 352–356.

[23] R.-C. Xu *et al.*, "Automatic target recognition based on super-resolution ISARimages," *Syst. Eng. Electron. Technol.*, vol. 28, no. 1, pp. 46–48, 2006.

[24] S. H. Park, J.-H. Jung, S.-H. Kim, and K.-T. Kim, "Efficient classificationof ISAR images using 2D Fourier transform and polar mapping," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 5, no. 3, pp. 1726–1736, Jul. 2015.

[25] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," 2017, *arXiv:1711.04043*.

[26] S. Thrun, "Lifelong learning algorithms," in *Learning to Learn*. Boston, MA, USA: Springer, 1998, pp. 181–209.

[27] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. 32nd Int. Conf. Mach. Learn. Workshop*, 2015.

[28] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. 30th Conf. Neural Inf. Process. Syst.*, 2016.

[29] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for fewshot learning," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017.

[30] Y. X. Wang, R. Girshick, M. Hebert, and B. Hariharan, "Low-shot learning from imaginary data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7278–7286.

[31] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2017, *arXiv:1711.04340*.

[32] L. Zhou *et al.*, "Learning to learn image classifiers with visual analogy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11497–11506.

[33] S. Flennerhag, A. A. Rusu, R. Pascanu, F. Visin, H. Yin, and R. Hadsell, "Meta-learning with warped gradient descent," 2019, *arXiv:1909.00025*.

[34] M. Guo *et al.*, "Neural graph matching networks for fewshot 3D action recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 653–669.

[35] A. Sperduti and A. Starita, "Supervised neural networks for the classification of structures," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 714–735, May 1997.

[36] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proc. Int. Joint Conf. Neural Netw.*, Dec. 2005, pp. 729–734.

[37] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[38] C. Gallicchio and A. Micheli, "Graph echo state networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2010, pp. 1–8.

[39] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[40] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.

[41] A. Micheli, "Neural network for graphs: A contextual constructive approach," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 498–511, Mar. 2009.

[42] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," in *Proc. 30th Conf. Neural Inf. Process. Syst.*, 2016, pp. 1993–2001.

[43] F. Sung, Y. Yang, L. Zhang, T. Xang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1199–1208.

[44] G. Hankinson, "Relational network brands: Towards a conceptual model of place brands," *J. Vacation Marketing*, vol. 10, no. 2, pp. 109–121, 2004.

[45] M. A. Rodriguez and J. Shinavier, "Exposing multi-relational networks to single-relational network analysis algorithms," *J. Informetrics*, vol. 4, no. 1, pp. 29–41, 2010.

[46] R. B. Palm, U. Paquet, and O. Winther, "Recurrent relational networks," 2017, *arXiv:1711.08028*.

[47] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[48] S. M. Alexander and D. Armitage, "A social relational network perspective for MPA science," *Conservation Lett.*, vol. 8, no. 1, pp. 1–13, 2015.

[49] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.

[50] V. C. Chen and M. Martorella, *Inverse Synthetic Aperture Radar*. Raleigh, NC, USA: Scitech Publishing, 2014.

[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-Cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 618–626.

[52] A. Lazarov and C. Minchev, "ISAR image recognition algorithm and neural network implementation," *Cybern. Inf. Technol.*, vol. 17, no. 4, pp. 183–199, 2017.

[53] Q. Zheng and J. Lafferty, "A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements," 2015, *arXiv:1506.06081*.

[54] C. Finn, P. Abbeel and S. Levine,"Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

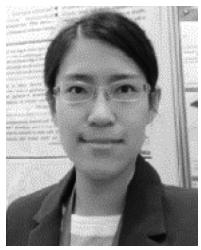
[55] 2018. [Online]. Available: https://www.fhr.fraunhofer.de/en/press-media/press-releases/2018/reentry_tiangong-1.html

Yun Zhang (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering and the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2000, 2003, and 2009, respectively.

She is currently a Professor with the Research Institute of Electronic Engineering Technology, HIT. Her research interests include SAR stereoscopic imaging stereo imaging and information acquisition.

Haoxuan Yuan (Student Member, IEEE) received the bachelor's degree in electrical engineering from Harbin Institute of Technology, Harbin, China, in 2019. He is currently working toward the Ph.D. degree in information and communication engineering with Harbin Institute of Technology, Harbin, China.

His research interests include stereo imaging and recognition of space targets.



Hongbo Li was born in Harbin, China. She received the M.S. degree in electronic information engineering from the Department of Communication and Information, Harbin Institute of Technology, Harbin, China, in 2009.

She is currently an Associate Researcher and Master's Supervisor with the School of Electronics and Information Engineering, Harbin Institute of Technology. Her research interests include target data processing technology, information counter measure technology, etc.

Jiaying Chen received the bachelor's degree in internet of things engineering from the School of Electrical and Information, Northeast Agricultural University, Harbin, China, in 2020. She is currently working toward the postgraduate degree in electronic information engineering from the School of Aeronautics and Astronautics, University of Electronic Science and Technology, Chendu, China.

Her research interests include target recognition and perception.

Muqun Niu was born Heilongjiang, China, in 1998. He received the B.S. degree in electronic and information engineering from the School of Electronic and Information Engineering, Harbin Institute of Technology (HIT), Harbin, China, in 2020. He is currently working toward the M.S. degree in information and communication engineering at the School of Electronic and Information Engineering, HIT.

His research interests include recognition for space targets, data fusion of heterogeneous sensors, and intelligent situational awareness.