

Domain Knowledge Powered Two-Stream Deep Network for Few-Shot SAR Vehicle Recognition

Linbin Zhang¹, Xiangguang Leng¹, *Member, IEEE*, Sijia Feng¹, Xiaojie Ma¹, Kefeng Ji¹, *Member, IEEE*, Gangyao Kuang, *Senior Member, IEEE*, and Li Liu², *Senior Member, IEEE*

Abstract—Synthetic aperture radar (SAR) target recognition faces the challenge that there are very little labeled data. Although few-shot learning methods are developed to extract more information from a small amount of labeled data to avoid overfitting problems, recent few-shot or limited-data SAR target recognition algorithms overlook the unique SAR imaging mechanism. Domain knowledge-powered two-stream deep network (DKTS-N) is proposed in this study, which incorporates SAR domain knowledge related to the azimuth angle, the amplitude, and the phase data of vehicles, making it a pioneering work in few-shot SAR vehicle recognition. The two-stream deep network, extracting the features of the entire image and image patches, is proposed for more effective use of the SAR domain knowledge. To measure the structural information distance between the global and local features of vehicles, the deep Earth mover's distance is improved to cope with the features from a two-stream deep network. Considering the sensitivity of the azimuth angle in SAR vehicle recognition, the nearest neighbor classifier replaces the structured fully connected layer for K -shot classification. All experiments are conducted under the configuration that the SRSIM and the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset work as a source and target task, respectively. Our proposed DKTS-N achieved 49.26% and 96.15% under ten-way one-shot and ten-way 25-shot, whose labeled samples are randomly selected from the training set. In standard operating condition (SOC) as well as three extended operating conditions (EOCs), DKTS-N demonstrated overwhelming advantages in accuracy and time consumption compared with other few-shot learning methods in K -shot recognition tasks.

Index Terms—Complex value information, feature extraction, few-shot learning, image classification, synthetic aperture radar (SAR) target recognition.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) automatic target recognition (ATR) has applicability in military and civilian

Manuscript received June 26, 2021; revised September 13, 2021; accepted September 20, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61872379 and Grant 62001480, and in part by Hunan Provincial Natural Science Foundation of China under Grant 2018JJ3613 and Grant 2021JJ40684. (*Corresponding author: Kefeng Ji.*)

Linbin Zhang, Xiangguang Leng, Sijia Feng, Xiaojie Ma, Kefeng Ji, and Gangyao Kuang are with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology, Changsha 410073, China (e-mail: zlbndt@163.com; luckight@163.com; fengsijia12@nudt.edu.cn; mxj286@foxmail.com; jikefeng@nudt.edu.cn; kuangyeats@hotmail.com).

Li Liu is with the College of System Engineering, National University of Defense Technology, Changsha 410073, China, and also with the Center for Machine Vision and Signal Analysis, University of Oulu, 90570 Oulu, Finland (e-mail: li.liu@oulu.fi).

Digital Object Identifier 10.1109/TGRS.2021.3116349

fields [1]. The situation that more and more scholars lay emphasis on the SAR ATR is of long-standing [2]–[8]. Great promotion to RGB imagery, brought about by deep learning [9], is enabled by the availability of massive amounts of RGB data. For instance, there are hundreds of millions of training images in ImageNet [10]. Although this ensures a promising future for SAR ATR, the application of deep learning in SAR ATR is not as smooth as it is in other problems in the field of computer vision [11]. The main reason for this is that deep learning requires a very large amount of data; otherwise, the classifier will be trapped into overfitting. Due to the flight track of carriers and the side view of radar, the vehicle's image will change sharply with the transformation of the depression angle and azimuth angle. As a result, SAR can only capture several samples at a restricted angle for the noncooperative vehicles, which are the enemy vehicles and cannot be put in the ideal positions or poses to get the SAR images. This situation leads to a few-shot environment for SAR vehicle recognition. Lack of training data limits the use of large-scale deep convolution neural networks (DCNNs) with powerful representation learning ability.

In recent years, some state-of-the-art networks, such as A-ConvNet [7] and data augmentation means [12]–[14], have been applied to SAR ATR. DCNNs are able to reduce the effect of speckle noise in SAR images [1] as well as avoid the trouble of manual notation of domain knowledge [7]. When there are a sufficient number of labeled samples, these methods are able to achieve significant recognition accuracy. However, when the number of training samples drops sharply, the performance of existing algorithms also becomes unsatisfactory, making it difficult to meet the requirements for real-world applications. In contrast, human vision systems can learn a visual concept quickly from very few samples [15]. Therefore, it is important for algorithms to be able to achieve this ability so that the cost of data labeling can be reduced. The use of few-shot learning purpose is growing in the fields of computer vision and pattern recognition. However, only a few methods [16]–[20], including transfer learning [21] and metric or loss based improvement, are used in SAR image interpretation. Unfortunately, these methods emphasize the improvements of models and networks but overlook the mechanism, which contains SAR domain knowledge, in SAR targets. SAR images are totally different from optical images in terms of imaging modalities, as shown in Fig. 1. Few-shot learning algorithms tend to extract more information from

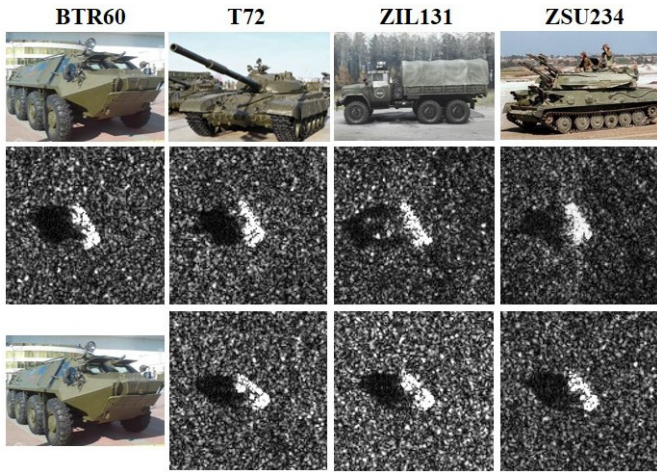


Fig. 1. First row: RGB images of BTR60, T72, ZIL131, and ZSU234, which are all armored vehicle categories from the MSTAR database. Second row: SAR images in the depression angle of 17° and azimuth angle of 122° in these four categories. Third row: BTR60 in an optical image and three SAR images in the depression angle of 17° and azimuth angle of 137° , 142° , 147° .

labeled samples, including expert knowledge, image features, and metric distances.

In this study, information about the phase data, and the azimuth angle, is used as SAR domain knowledge to overcome the problem of a lack of labeled data. In essence, the amplitude and phase information forms the complex images of SAR because of the coherence of microwave imaging. Most CNN approaches designed for SAR image target classification only take the amplitude information into consideration and discard the phase information. However, the phase information is believed to offer complementary information in high-resolution single-channel SAR imagery [22]. This study adopts a data augmentation method on generating in-phase and quadrature the amplitude data within the cos phase data and sin phase data [13]. Then, a single-channel grayscale SAR image can be extended to a three-channel RGB image so that both amplitude information and phase information are considered, as shown in Fig. 2. In addition, azimuth angles describe the different backscattering information of vehicles in such a position, so many algorithms make use of azimuth angles in data enrichment or feature joint learning [12], [14], [23]–[26]. In this study, the azimuth angle of each vehicle is rotated in a counterclockwise direction, called azimuth angle normalization, to ensure that the image area of different vehicles in SAR image is a one-to-one match to each other, which can be considered as target-level image registration. Introducing complex data allows the enhancement of local information to a single vehicle, and azimuth angle normalization helps to achieve the image patches in a fixed location, which corresponds to the same part of the vehicles. These enhanced local representations supply discriminative and transferable information across categories, which are important for SAR vehicle recognition in few-shot situations.

In this study, a two-stream deep network, as shown in Fig. 4, is proposed to extract the global features from the entire vehicle image and the local features, which owns the index and order of positions, from image patches. Sharing the

same structure, network, and parameters, the two-stream deep network has two different inputs and outputs to each stream. To measure the distance of local and global features among different images, deep Earth mover's distance (EMD) [27], [28] is adopted and improved for SAR vehicle targets. The positions where the interest targets appear in the optical image are random in the image, but vehicles can be easily located at the center of the SAR image. Thus, given a group of local and global features extracted by two SAR images, the inherent structural similarity of these two SAR images can be computed by EMD [29]. In SAR images, the positions of shadows and bright spots of SAR targets with different azimuth angles are different. If K -shot SAR images in one category are represented by one prototype, the global and local features are not as intuitive as a single image. This will influence the calculation of EMD and the prototype after structured fully connected (SFC) layers will be hard to distinguish with other categories. Hence, in this study, we design nearest neighbor (NN) classifiers to replace the SFC layer in K -shot settings in SAR vehicle target recognition, which fully considers the sensitivity of azimuth angle. In particular, when the number of training images increases, shadows of vehicle targets are in more and more different directions in the images, leading to the fuzzy features to such category. In addition, the update of SFC in [28] is more time-consuming and less accurate in K -shot SAR vehicle target recognition, compared to the NN classifier.

According to the baseline in few-shot learning, categories in source tasks will not appear in the target tasks. The number of labeled samples in the source tasks is sufficient, whereas the labeled samples in the target tasks are few-shot. To ensure the effectiveness of our domain knowledge powered two-stream deep network (DKTS-N), target tasks are sampled from the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset with standard operating condition (SOC) and three extended operating conditions (EOCs) [30], EOC1 (LARGE DEPRESSION VARIATION), EOC2/C (CONFIGURATION VARIATION), and EOC2/V (VERSION VARIATION), whereas source tasks are from the Sarsim dataset [31], which consists of the simulated data with the same imaging mode and parameters as the MSTAR. Our goal is that the few-shot algorithms can learn the transfer information from the simulated dataset and generalize it to the MSTAR dataset. Extensive experiments were conducted to demonstrate the effectiveness of our algorithm. The three main contributions of our study can be summarized as follows.

- 1) We explore the framework of SAR image data based on electromagnetic simulation to support few-shot SAR vehicle recognition. A baseline is also proposed to use the public Sarsim dataset as source tasks and the public MSTAR dataset as target tasks in few-shot SAR target recognition. This configuration is the most challenging in the current research of SAR target small sample recognition and it can be easily generalized to other SAR dataset in the future as well.
- 2) A two-stream deep network is proposed to utilize the SAR domain knowledge. One stream in the deep network aims to extract the global features of the entire

image and the other stream extracts the local features of the vehicles' SAR image patches. In addition, EMD is used to calculate the structural information between global and local representations. The optimal flow of different representations is restricted by the proposed coefficient matrix. The NN classifier replaces the SFC layers in K -shot scenery.

- 3) Domain knowledge about the phase data and the azimuth angle normalization is adopted and enhanced the few-shot labeled data. With the augmented SAR data, our proposed DKTS-N algorithm achieves 49.26% and 96.15% under ten-way one-shot and ten-way 25-shot in the MSTAR SOC. A plenty of contrast experiments and ablation experiments show that our DKTS-N is superior to other few-shot algorithms in recognition rate and the augmentation methods for few-shot SAR vehicle recognition are effective.

The remainder of this article is composed of five sections. In Section II, the related work about few-shot SAR target recognition is introduced, SAR data augmentation and EMD. In Section III-A, the effect of azimuth angle normalization is described. The use of complex information to generate RGB SAR images is presented in Section III-B. Then, the problem settings of few-shot SAR vehicle target recognition and the details of DKTS-N are introduced in IV, including one-shot and K -shot situations. Experimental results and implementation details are summarized and presented in Section V. In the end, Section VI concludes this article and designs the future work.

II. RELATED WORK

A. Few-Shot SAR Target Recognition

1) *Few-Shot Learning*: Few-shot learning is studying from limited supervised information to get the hang of the task. It has shown promise for learning and adapting from a few samples in source tasks and avoiding overfitting in the target tasks. There are essentially two ways to solve few-shot learning problems: optimization-based and metric-based. Optimization-based algorithms construct novel optimization functions or better initialization of training models to improve the rapid adaptability to new tasks is also a common solution in few-shot learning methods. For instance, MAML [32] have shown impressive results on the few-shot learning datasets MiniImageNet [33] and Ominiglot [34] at the early time of the development in few-shot learning based on CNNs [35]. The metric learning method does not focus on fine-tuning the support set (few labeled samples) but classifies by judging the distance between the query set image and the support set image, such as matching networks [36], prototypical networks [37], and deep nearest neighbor neural network (DN4) [38]. In addition, DN4 has been improved to D2N4 to cope with the few-shot recognition mission in space targets [39].

2) *Few-Shot SAR Target Recognition*: SAR target recognition is increasingly used in both civil and military arenas. Of the many methods proposed [2], [5]–[8], [25], DCNNs have achieved 99.13% [7] in the MSTAR public dataset. However,

for data collection of noncooperative targets, it is common to face the challenge of limited labeled SAR data. Some scholars have focused on solving the problem of few-shot SAR target recognition [16]–[19], [40]–[42]. Siamese neural network, whose target category was outputted by the classifier [18] and not by the similarity discriminator [43], was improved to solve the challenges of few-shot SAR target recognition. Instead of cross-entropy loss, the triple loss was introduced to the deep learning framework for SAR image recognition with the experiment configuration about only one few-shot class and nine classes with sufficient data [16]. A meta-learning framework named MSAR [41], consisting of a meta-learner and a base learner, can learn a good initialization as well as a proper update strategy. A hybrid inference network (HIN) [42] adopts the inductive inference and the transductive inference to classify the samples in the embedding space. These two methods divide the MSTAR dataset into query set and support set and the performance mentioned in [41] and [42] cannot be reflected on the whole MSTAR dataset. These algorithms are simply improved from the few-shot learning algorithm in optical images without considering the characteristics of SAR targets.

B. SAR Data Augmentation

1) *Complex SAR Target Classification*: In coherent microwave imaging, each pixel in an SAR image contains not only the grayscale information reflecting the surface microwave reflection intensity but also the phase data. As we know, complex data are widely used in polarimetric SAR (polSAR) [44], [45]. A complex-valued convolutional neural network (CV-CNN) [45] had higher classification accuracy than the real neural network with the same network structure and parameter settings. When it comes to the single-channel SAR images, the methods [13], [46] are designed for the MSTAR dataset and significantly improve the classification accuracy. What is more, Leng *et al.* [47] proposed a new ship detection method based on complex signal kurtosis (CSK) in single-channel SAR imagery. In this study, SAR targets' phase data were used to generate romantic SAR RGB images as one method of augmentation.

2) *SAR Target Recognition Based on Azimuth Angle*: In [14] and [23]–[26], azimuth angle has been used as a sequence of information to improve the effect of deep learning. The vehicle target at a specific azimuth can show the electromagnetic scattering characteristics of the vehicle body structure and details of components at this azimuth angle [48]. Targets with similar azimuth angles are put into a group for learning. Hence, algorithms are expected to find the common features of the vehicle targets in the azimuth angle sequence group. There are also several data expansion techniques involving azimuth angles [12]. However, in the real world, there are few labeled samples in random azimuth angles, and thus, it is difficult to use the above-mentioned approaches. In this study, through azimuth angle normalization, these targets are in the same direction and location in SAR images, making it easier to extract and compare local representations from the two-stream deep network.

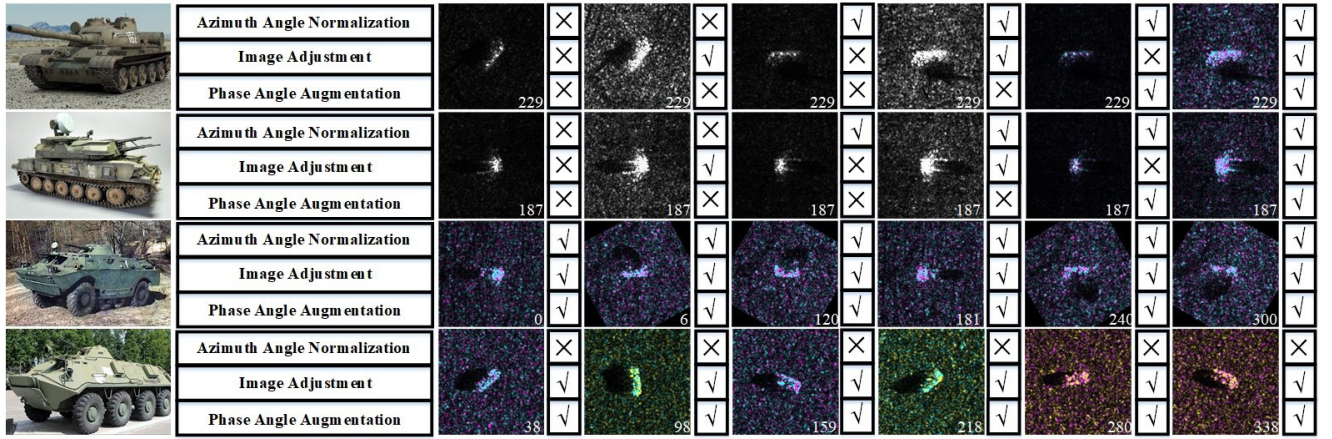


Fig. 2. RGB images and SAR images in different processing of the armored vehicles. From the first row to the fourth row, the images describe the vehicles in different categories, T62, ZSU234, BRDM2, and BTR60. The azimuth angle of each SAR target is shown in the bottom-right corner of the image. In addition, the three symbols of tick or cross from top to bottom, on the right of images, denote whether the SAR image is under conducted the image processing of azimuth angle normalization, image adjustment, and phase angle augmentation.

C. Earth Mover's Distance

The EMD metric is able to evaluate the dissimilarity between two multidimensional distributions in some feature space. It is also used to find the optimal matching flows to the transportation problem from linear optimization. In 2000, it was used to measure the distance between color and texture in image retrieval [29]. Since then, EMD has been applied in many fields, such as multiobject tracking [49], document classification [50], and visual tracking [51]. In [52], EMD was used to determine the similarity of two graphs, which are represented as a set of vectors corresponding to the vertices in gesture recognition. Li [53] used a tensor-SIFT-based EMD with kernel density for distribution modeling to tackle the problem of contour tracking in adverse conditions. EMD is also used in SAR image segmentation [54] and farmland detection [55] by some scholars. In 2020, it was used to solve recognition problems in few-shot scenery [27], [28], applying the implicit function theorem [56], [57] to form the Jacobian matrix. Inspired by the cross-reference mechanism in [28], parameter weights were adjusted from a proposed coefficient matrix to avoid the flow match between global and local representation.

III. SAR DOMAIN KNOWLEDGE

In this section, two data augmentation methods for SAR vehicle images are discussed. Both methods are designed for extracting more useful features from SAR data through the two-stream deep network. Azimuth angle normalization is mainly for target-level image registration, through which the location of local representations can be fixed, so that the learning processing is more specific. Transforming grayscale SAR images to RGB images with phase data enhances the details of local presentations. These two methods also enrich the experience of training data to solve the problem of few-shot recognition.

A. Azimuth Angle Normalization

In the background of SAR imaging to interested targets, such as the MSTAR dataset, airborne SAR sends and receives

radar signals through a side view with a depression angle and a height. Because targets remain on the ground with different azimuth angles, it is obvious that, in the original SAR image, shadows in every image are on one side of the vehicles. Due to the side-looking imaging mode of SAR, some areas of targets and their background lose their echo signal because it may be blocked by the turret and the targets themselves. Furthermore, attributed scattering centers [58]–[60], which reflect the structure of targets, also change as the azimuth angles change. Moreover, shadows of SAR targets also show the height, length, and width attributes of the vehicle. For example, in certain azimuth angles, the main gun of the tanks (T62 and T72) appears in the shadow, and the contour of the tractor operator cabin can be observed in the shadows as well, as shown in the top row in Fig. 2. Hence, the local representations of targets are significant enough to distinguish their labels from other categories. This expert knowledge helps to avoid overfitting and increases the recognition rate in the few-shot scenery.

To directly compare the structure of global and local representations of targets, azimuth angle normalization is proposed in this study as one method of SAR data augmentation. Rotate the image by the target's azimuth angle in the counterclockwise direction and then reduce the image size the same as the initial image in the center. Next, the front and rear of the vehicles toward the right and left, individually, which like the images with the click in azimuth angle normalization in Fig. 2. This data augmentation method adds prior knowledge to the data without more consumption of network structure or model capacity and it can also be used in the situation where only one-shot sample is given in each category.

B. Phase Data Augmentation

Most current single-channel SAR target recognition methods based on deep learning use only the amplitude information of the SAR image but ignore the phase information. The reason why these methods discard the phase information is that for low-resolution SAR sensor, the wavelength is much less than

resolution. However, in a statistical sense, the phase information in images from high-resolution SAR sensor is no longer completely random. For high-resolution single-channel SAR images, the phase information contains additional information, which is able to improve the recognition rate [22].

The complex-valued SAR image is available as follows:

$$Z(u, v) = \text{Re}\{z(u, v)\} + j\text{Im}\{z(u, v)\} \quad (1)$$

where $\text{Re}\{z(u, v)\}$ is the real part, $\text{Im}\{z(u, v)\}$ is the imaginary part of the complex-valued SAR image, and j is the imaginary unit that $j^2 = -1$. Note that (u, v) represents the complex-valued plane. Most approaches for SAR target recognition only use the amplitude data, i.e.,

$$|z(u, v)| = \sqrt{\text{Re}^2\{z(u, v)\} + \text{Im}^2\{z(u, v)\}}. \quad (2)$$

In this study, our proposed augmented RGB SAR image consists of three-channel grayscale image $|z(u, v)|$, $\text{Re}\{z(u, v)\}$, and $\text{Im}\{z(u, v)\}$ as follows:

$$Z(u, v) = [\text{Im}\{z(u, v)\} \text{Re}\{z(u, v)\} |z(u, v)|]. \quad (3)$$

If the three channels of RGB swap with each other in locations, the RGB image will be in a different style. However, it appears the same to the feature extractor network even if it is in various colors, as shown in the bottom row in Fig. 2.

In Fig. 3, the first and the second columns are RGB and SAR images of T72, in different parameter ranges of contrast stretch, while the others are D7. With respect to the SAR images, the first and third rows correspond to $\text{Re}\{z(u, v)\}$ and $\text{Im}\{z(u, v)\}$, respectively. It is obvious that the $Z(u, v)$ and $|z(u, v)|$ images differ only in the color of scattering points. In addition, the real part $\text{Re}\{z(u, v)\}$ and imaginary part $\text{Im}\{z(u, v)\}$ of complex data are the downsample of $|z(u, v)|$ in the second and fourth rows. In the data from high-resolution SAR sensors, such as the MSTAR dataset, the phase data have certain extra information in a statistical sense. Thus, the bright pixels, which are high energy points, in the real part and imaginary part are different. The results of the ablation experiments demonstrate that the recognition of complex RGB SAR images is higher than that of regular SAR images.

IV. STRUCTURE OF THE DKTS-N

To solve the problem of few-shot SAR vehicle target recognition using the aforementioned data augmentation methods, the DKTS-N framework, composed of SAR domain knowledge augmentation, the two-stream deep network, and the EMD module, is proposed. As shown in Fig. 4, the proposed approach takes one support image and one query image and inserts them into the DKTS-N. First, the global and local features are obtained through a deep embedding module. Subsequently, the structure distance of these features is measured by the EMD, and eventually, the distance value is output. In this section, improvements and the reasons behind these improvements reason are introduced. Although the optimization process and end-to-end training have been shown in [27] and [28], they will not be further discussed here.

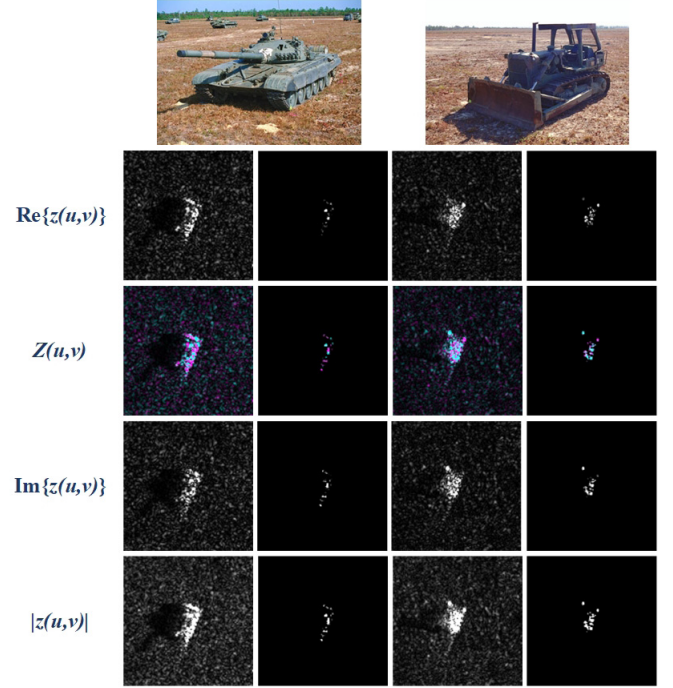


Fig. 3. Two examples of the complex data in SAR vehicle images (T72 and D7).

A. Problem Setting

Before discussing our algorithm in detail, the terminology used in few-shot SAR target recognition should be defined. N and K are the number of few-shot categories and the number of labeled data in each category, respectively. N -way K -shot is the basic set in few-shot learning and K is usually 1 or 5. Define $N \times K = s$. Then, assume that there is a large labeled dataset, which in this study is the SARSIM dataset, from which it is feasible to sample the support set $\mathcal{S}_{\text{sou}} = \{(F_m(x, y), l_m) | m = 1, 2, \dots, s, l_m \in C_{\text{sou}}\}$ and query set $\mathcal{Q}_{\text{sou}} = \{(F_n(x, y), l_n) | n = s + 1, \dots, p, l_n \in C_{\text{sou}}\}$. $F_m(x, y)$ indicates the image in the support set and $F_n(x, y)$ indicates the image in the query set. p indicates the sum number of samples in the SARSIM datasets. C_{sou} is the category set of the labeled dataset. l is the corresponding label of the sample $F(x, y)$. Our goal is to use the simulated data as source tasks and test the model in target tasks, which can also be defined as the support set $\mathcal{S}_{\text{tar}} = \{(F_m^*(x, y), l_m^*) | m = 1, \dots, s, l_m^* \in C_{\text{tar}}\}$ and query set $\mathcal{Q}_{\text{tar}} = \{(F_n^*(x, y), l_n^*) | n = s + 1, \dots, q, l_n^* \in C_{\text{tar}}\}$. $C_{\text{sou}} \cap C_{\text{tar}} = \emptyset$. q indicates the sum number of samples in the MSTAR SOC or EOCs datasets. At training time, all the data are sampled from the SARSIM dataset, and learning supervision is provided by the ground truth label of samples in the query sets. At inference time, tasks are sampled from the MSTAR dataset for evaluation, and their mean accuracy and variance are recorded. In our few-shot classification experiments, ten-way K -shot ($K = 1, 2, 5, 10, 20$, and 25) is set in the MSTAR SOC, and four-way K -shot ($K = 1, 2, 5, 10$, and 25) settings are selected in the MSTAR EOC1 (LARGE DEPRESSION VARIATION), EOC2/C (CONFIGURATION VARIATION), and EOC2/V (VERSION VARIATION).

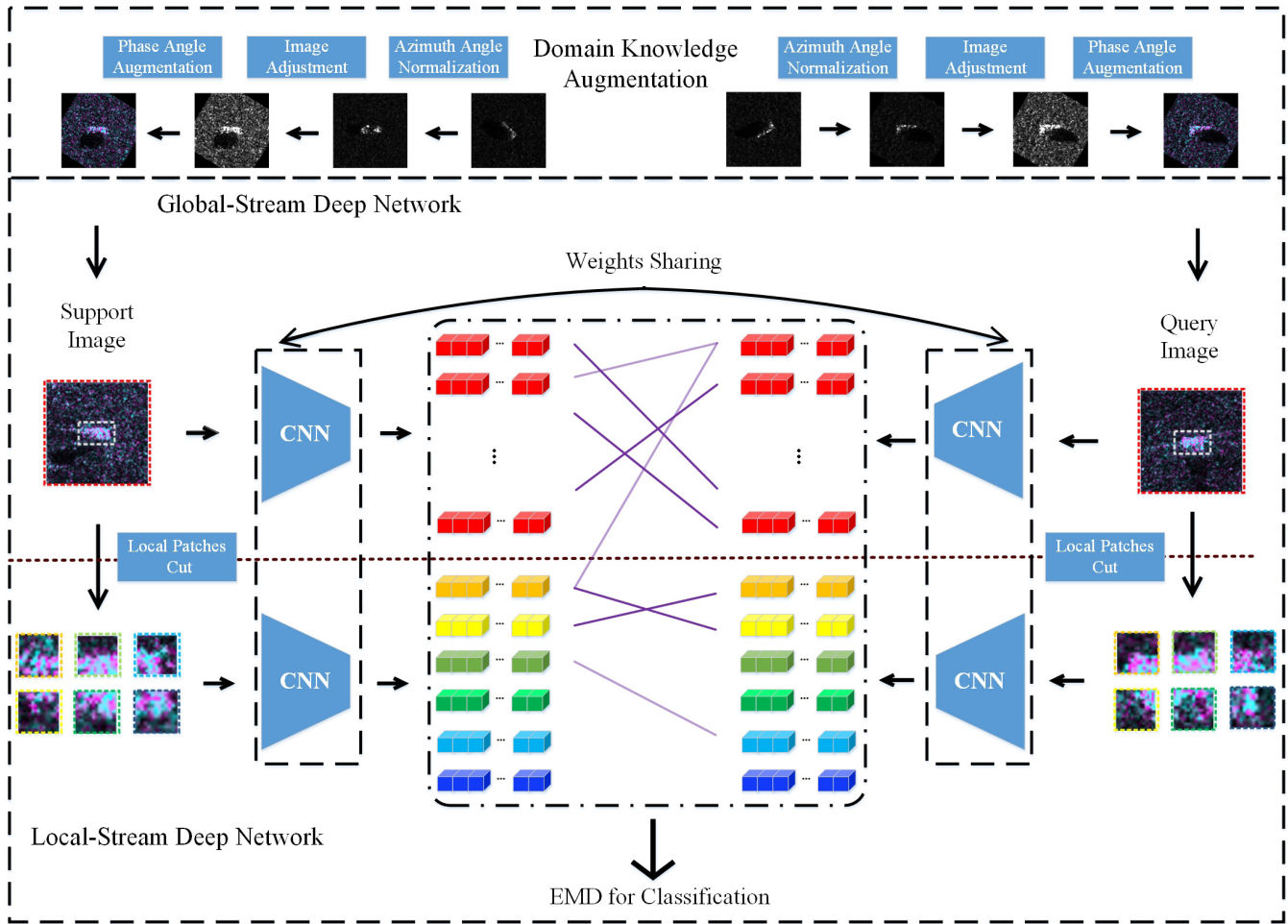


Fig. 4. DKTS-N includes the augmentation with SAR domain knowledge and the EMD based on global and local features of one-shot SAR vehicle target recognition. When a pair of SAR vehicle images are given, the CNNs, sharing the same weights, are used to extract global and local feature vectors of both images. After that, the weights of each vector are generated from the above-mentioned feature vectors based on a cross-reference mechanism. Subsequently, EMD is utilized to evaluate the cost of transporting the features of the support image to the features of the query image. Finally, the distance of two images for classification is calculated according to the optimal matching flows and costs.

B. Two-Stream Deep Network

The two-stream deep network aims to extract discriminative global and local features by four fully convolutional blocks [61], which are similar to the four conventional blocks in DeepEMD [28]. Each conventional block contains a 3×3 convolution with different numbers of filters, 2×2 max-pooling layers, a batch normalization layer, and a leaky RELU (0.5) nonlinearity layer. The difference is that the number of filters in each block is changed to fit the mission in few-shot SAR vehicle targets recognition but not the optical images in the experiments of [28].

The image is cut into six patches, and each patch is put into the embedding module to generate six local representations $[\mathbf{local}_1, \dots, \mathbf{local}_6]$, $\mathbf{local}_i \in \mathbb{R}^d$ by one stream, which is shown in dark yellow, light yellow, dark green, light green, dark blue and light blue in Fig. 4. The image is also sent to achieve eight global representations $[\mathbf{global}_1, \dots, \mathbf{global}_8]$, $\mathbf{global}_j \in \mathbb{R}^d$ by the other stream, in the gradient red colors show in Fig. 4. However, this image is completely different from the form of the feature map

in DeepEMD [28]. The support image and query image are transmitted into the network, which shares the same weight for extracting the global and local representations. Hence, these two sets of representations are able to measure the optimal matching cost of the dissimilarity of two images. It is worth noting that armored vehicles are put in the center of SAR images, so the six image patches are located so as to cover the target. In addition, the structure of vehicles is symmetric about its centerline, which indicates that the local representations in dark color and light color in Fig. 4 may share a close distance. Due to the azimuth angle normalization, it is convenient to measure the structural distance of representations with EMD. Because the shadow of the target may occur in a certain direction according to the azimuth angle, both the global and local representations should be considered for classification. In [28], feature maps contain too many representations, which will prove highly time-consuming in calculating the weight matrix of vectors and the optimal flow cost matrix. Our DKTS-N extracts further information in SAR vehicle targets with fewer representations with shorter dimensions and is less time-consuming than DeepEMD.

C. EMD for Classification

The EMD was first proposed for image retrieval by Rubner *et al.* [29]. It is based on a solution to the transportation problem [62] from linear optimization. In essence, this distance is to measure the structural information of representations from two sets. In 2020, its use was proposed in few-shot learning [28], with the name DeepEMD. In addition, in [27], DeepEMD-FCN, DeepEMD-Grid, and DeepEMD-Sampling were proposed for slightly higher recognition accuracy in optical few-shot datasets, but the novel structure of DeepEMD requires much more video memory and more calculation time.

In the background of few-shot SAR vehicle target recognition, suppose that the representations of support image and query image are $[\mathbf{local}_1, \dots, \mathbf{local}_6, \mathbf{global}_1, \dots, \mathbf{global}_8]$ and $[\mathbf{local}_1^*, \dots, \mathbf{local}_6^*, \mathbf{global}_1^*, \dots, \mathbf{global}_8^*]$ individually. For the sake of simplicity, $\mathcal{S} = \{\mathbf{s}_i | 1 \leq i \leq 14\}$ indicates the global and local representations of the support image and $\mathcal{Q} = \{\mathbf{s}_j^* | 1 \leq j \leq 14\}$ indicates the global and local representations of the query image. Each representation owns its weight w_i for the support image or w_j^* for the query image. The cost per unit transported from w_i to w_j^* is defined by c_{ij} , and the number of units transported is denoted by t_{ij} . It is required that we find the optimal flow $\mathcal{T} = \{t_{ij} | 1 \leq i \leq 14, 1 \leq j \leq 14\}$ at the least cost from transporting w_i to w_j^*

$$\begin{aligned} \min_{t_{ij}} \quad & \sum_{i=1}^{14} \sum_{j=1}^{14} \alpha_{ij} c_{ij} t_{ij} \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^{14} t_{ij} = w_i, & 1 \leq i \leq 14 \\ \sum_{j=1}^{14} t_{ij} = w_j^*, & 1 \leq j \leq 14 \\ t_{ij} \geq 0, & 1 \leq i, j \leq 14 \end{cases} \end{aligned} \quad (4)$$

$$\alpha_{ij} = \begin{cases} 1, & 1 \leq i, j \leq 6, 7 \leq i, j \leq 14 \\ 10, & \text{others.} \end{cases} \quad (5)$$

To ensure that the weights of global representations in support image would better be transported to the weights of global representations in the query image and to ensure that the weights of local representations in the support image would be better transported to the weights of local representations in the query image, α_{ij} is defined as an attention mechanism to limit the transport from weights of local representations in the support image to weights of global representations in the query image. This improvement is a good fit for the SAR vehicle target situation because the structures of these targets have a connection to each other and are not like the randomness of interested targets in optical images. c_{ij} , which indicates the cost per unit transported from w_i to w_j^* , is defined by the representations \mathbf{s}_i and \mathbf{s}_j^* of the support image and the query image

$$c_{ij} = 1 - \frac{\mathbf{s}_i^T \mathbf{s}_j^*}{\|\mathbf{s}_i\| \|\mathbf{s}_j^*\|}. \quad (6)$$

The weights of representations are also important to the EMD distance. In few-shot SAR vehicle target classification tasks, representations for recognition often contain high-level semantic information, so it is not unusual to find that there are shadows or strong scattering centers in the background regions

of target objects. Sometimes, it is hard to decide whether the shadow or the component parts of the vehicle determine the final category of the targets. In the similar situation of an optical image in the generation of weights, we adopt the regular rules of the cross-reference mechanism that uses the dot product between a node representation and the average representation in the other structure to generate the weights w_i and w_j^* , which is proposed in [27]

$$\hat{w}_i = \max \left\{ \mathbf{s}_i^T \cdot \frac{\sum_{j=1}^{14} \mathbf{s}_j^*}{14}, 0 \right\} \cdot \frac{14}{\sum_{j=1}^{14} s_j} \quad (7)$$

$$\hat{w}_j^* = \max \left\{ \mathbf{s}_j^{*T} \cdot \frac{\sum_{i=1}^{14} \mathbf{s}_i}{14}, 0 \right\} \cdot \frac{14}{\sum_{i=1}^{14} s_i^*}. \quad (8)$$

After considering all the variables in the optimal matching flows, the similarity scores between images can be computed as follows:

$$\text{Score}(\mathbf{s}, \mathbf{s}^*) = \sum_{i=1}^{14} \sum_{j=1}^{14} \alpha_{ij} \cdot (1 - c_{ij}) \cdot t_{ij}. \quad (9)$$

D. NN for K-Shot Classification

In [27] and [28], an SFC layer is proposed to learn a prototype feature map generated by a dummy image for each class after several updates. This layer is effective in optical images according to the analysis and results in [27] and [28]. However, when it comes to SAR vehicle target recognition, this layer is not as effective as in optical images, and it will perform even worse when the number of shots added is 10 or 20. Due to the sensitivity of SAR vehicle images to azimuth angles, it is ill-suited to use the prototype representations to describe a category. To be specific, if there are only two samples with the 90° and 180° value of azimuth angles, the prototypical representations of these two samples will be weakened because the shadow and scattering points in these two samples are completely different. Based on the properties in SAR vehicle images, the NN classifier is proposed as a substitute for SFC. The structural information of global and local representations will be closer if the changed azimuth angle or depression angle is small among one category. Therefore, to calculate the NN easily, every K -shot image from the support set in one category is first compared to a query image to find the minimal cost in that category. Next, the representations of the nearest samples based on DKTS-N distance in each category are measured again with the query image to determine its final label. With this adjustment, much time is saved in few-shot SAR vehicle target recognition for leaving out the updates in SFC. The accuracy rate soars with this improvement, which takes the azimuth angle into consideration.

V. EXPERIMENTS

To verify the effectiveness of DKTS-N for SAR vehicle targets, experiments were extensively conducted in few-shot ($k \leq 5$) and limited-labeled ($k \geq 5$) situations. In terms of SAR vehicle target recognition, the public simulated SARSIM dataset was recognized as source tasks, and the public MSTAR dataset was recognized as target tasks. Contrast experiments with state-of-the-art few-shot learning approaches were conducted. In addition, ablation experiments with different parameters or experiment settings are also involved in our work.

All experiments were run on a PC with an Intel single-core i9 CPU, four Nvidia GTX-2080 Ti GPUs (12 GB VRAM each), and 128-GB RAM. The PC operating system was Ubuntu 20.04. All experiments were conducted using the Python language on the PyTorch deep learning framework and CUDA 10.2 toolkit.

A. DataSets

1) *SARSIM*: The SARSIM dataset consists of simulated SAR images of vehicles. They are simulated in the same imaging conditions as *MSTAR*, according to the description in [31]. This dataset contains seven different vehicles with two different objects in each category (truck 2107 and 2096, car Toyota and Peugeot607, motorbike 3651_Suzuki and 3972, bus 30726 and 55473, tank 65047 and 86347, bulldozer 13013 and 8020, and humvee 3663 and 9657). The simulated data can be downloaded from the Web. Each target is simulated for every 2° azimuth angle object rotation at seven different depression angles (15° , 17° , 25° , 30° , 35° , 40° , and 45°). Each object has been simulated with three statistically generated background clutter types, grass, roads, and medium, which is the mean value of the grass and roads. Because there are actually 14 categories in the SARSIM dataset, it is feasible to sample ten-way classification as SOC and four-way classification as EOCs in the *MSTAR* dataset. This is an ideal way to confirm the effectiveness of few-shot learning methods in SAR target recognition. The details of the SARSIM dataset are shown in Table I.

2) *MSTAR*: In recent years, many algorithms were created and compared on the SAR images in the *MSTAR* dataset [63]. The X-band imaging radar works in the HH polarization mode, and the obtained image size is 128×128 pixels with a resolution of $0.3 \text{ m} \times 0.3 \text{ m}$. The published *MSTAR* dataset, which was supported by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL), contains ten types of Soviet military vehicle targets (T-72 and T-62, tanks; BTR-60, BTR-70, BMP-2, and BRDM-2, armored vehicles; ZIL-131, Military truck; ZSU-234, Self-propelled artillery; 2S1, Self-propelled howitzer; and D7, Bulldozer). The specific training and testing samples are shown in Table II. SOC, samples in a depression angle of 17° were for training and 15° were for testing, was recognized as the standard dataset for our ablation experiments. Few-shot settings of ten-way K -shot were conducted in SOC. To verify the satisfactory generalization performance and robustness of DKTS-N, experiments in EOC1, EOC2/C, and EOC2/V were conducted in four-way K -shot settings.

The EOC1 experiment involved four targets (2S1, BRDM-2, T-72, and ZSU-234). The depression angles of the support set and query set were 17° and 30° , respectively, which reflects a more obvious change than the 17° and 15° in SOC. The settings of the samples are shown in Table II. The EOC2/C experiment was about the configuration variation, where the targets in the support set and query set were different in components such as reactive armor and auxiliary fuel barrel. The support set contained four targets (BMP-2, BRDM-2, BTR-70, and T-72), whereas the query set contained only the T-72 target with five configuration variations. The EOC2/V

TABLE I
SARSIM DATASET WITH DIFFERENT BACKGROUND
GRASS\MEDIUM\ROADS

Target	15	17	25	30	35	40	45
truck 2107	72	72	72	72	72	72	72
truck 2096	72	72	72	72	72	72	72
car Toyota	72	72	72	72	72	72	72
car Peugeot 607	72	72	72	72	72	72	72
motorbike 3651	72	72	72	72	72	72	72
motorbike 3972	72	72	72	72	72	72	72
bus 30726	72	72	72	72	72	72	72
bus 55473	72	72	72	72	72	72	72
tank 86347	72	72	72	72	72	72	72
tank 65047	72	72	72	72	72	72	72
bulldozer 13013	72	72	72	72	72	72	72
bulldozer 8020	72	72	72	72	72	72	72
humvee 3663	72	72	72	72	72	72	72
humvee 9657	72	72	72	72	72	72	72

TABLE II
SUPPORT SET AND QUERY SET FROM *MSTAR* SOC

SOC	Support Set	Query Set
2S1	299	274
BMP2	233	196
BRDM2	298	274
BTR60	256	195
BTR70	233	196
D7	299	274
T62	299	274
T72	232	196
ZIL131	299	274
ZSU234	299	274

corresponds to the target version variation, which means that after some armored vehicles are finalized, they will be upgraded, such as adding the state-of-the-art reactive armor or replacing the main gun with a larger caliber. Similarly, they will also be downgraded or weakened for export trade, which is the variation of weapon version. The EOC2/V and EOC2/C experiments share the exact same support set, but the query set of EOC2/V includes five versions of T-72 and two versions of BMP-2. The settings of all three EOCs are shown in Table III.

The target tasks in the experiment settings of N -way K -shot were sampled from *MSTAR* SOC or EOCs. N -way indicates all ten categories in SOC or four categories in EOCs, with K labeled samples in each category. These labeled images are sampled from the support sets shown in Tables II and III. Samples in the query sets shown in Tables II and III are for evaluation. During the process of the few-shot learning in source tasks, all the labels of samples in the SARSIM dataset are known and are divided into support set and query set, which is similar to the target tasks. Due to the lack of phase data in the SARSIM dataset, only grayscale images were obtained in the few-shot training process.

B. Experimental Results

In order to make the samples in the SARSIM dataset and the *MSTAR* dataset go through the same preprocessing steps, the samples in the SARSIM database are also experienced

TABLE III
SUPPORT SET AND QUERY SET FROM MSTAR EOCs

Target	Support Set	Target(EOC1)	Query Set
2S1	299	2S1(b01)	288
BRDM2	298	BRDM2(E71)	287
T72	691	T72(A64)	288
ZSU234	299	ZSU234(d08)	288
Target	Support Set	Target(EOC2/C)	Query Set
BMP2(9563)	233	T72(S7)	419
BRDM2(E71)	298	T72(A32)	572
BTR70(c71)	233	T72(A62)	573
T72(SN132)	232	T72(A63)	573
		T72(A64)	573
Target	Support Set	Target(EOC2/V)	Query Set
BMP2(9563)	233	T72(SN812)	426
		T72(A04)	573
BRDM2(E71)	298	T72(A05)	573
		T72(A07)	573
BTR70(c71)	233	T72(A10)	567
		BMP2(9566)	428
T72(SN132)	232	BMP2(C21)	429

TABLE IV
RESULTS OF SOC AMONG ALGORITHMS IN LIMITED DATA

Algorithms	10 way 10 shot	10 way 20 shot	10 way 30 shot
A-ConvNet [7]	44.5	70.4	86.7
VGG16 [64]	34.9	66.4	76.6
Resnet-18 [65]	27.7	68.5	66.1
ESENet [66]	54.4	79.4	89.8
ShuffleNetV2 [67]	43.2	73.9	86.9
SiameseNet+ [68]	67.7	85.8	92.6
Fast Inference [18]	74.7	78.7	92.6
Our DKTS-N	87.0	93.0	97.1

azimuth angle normalization and image adjustment before training. However, lack of phase data in the Sarsim dataset, the grayscale images are prepared for source task training. When it comes to the SOC dataset, the few-shot settings of ten-way K -shot were selected, and in the EOCs, four-way K -shot classification tasks were performed. The targets were 128×128 with all the data augmentation in this study in both the support set and the query set. The phase data information and azimuth angles are known at first. The numbers of filters in the four convolution blocks are 32, 64, 96, and 128, and in the ablation experiments, other numbers were displayed to show that the selected numbers are suitable for few-shot SAR vehicle recognition.

1) *Experiments in SOC*: In Table IV, other algorithms [7], [64]–[68] are compared to DKTS-N. Ten-way K -shot involved in the experiments were randomly selected with a 17° depression angle from the SOC dataset. After conducting 1000 experiments, the average recognition rate was recorded. It is obvious that as the number of training samples increases, the recognition rate rises dramatically. The recognition rate of ten-way 30-shot is nearly twice as much as that of ten-way ten-shot, which also confirms that these approaches perform poorly when the number of training samples is limited. Siamese Net+ [68] and the fast inference network [18] method are higher than the above-mentioned methods, and their results are

over 90% in ten-way 30-shot. Our DKTS-N achieved 84.6% in ten-way ten-shot, which demonstrates that it can achieve excellent performance even when the number of training samples is limited. When the number of training samples increases to 30-shot per category, DKTS-N obtains over 97% in the recognition task.

Other few-shot learning methods are compared in Table V, with the experiment settings of ten-way one-shot, ten-way two-shot, ten-way five-shot, ten-way ten-shot, and ten-way 25-shot. In the situation of SOC, there are ten categories, while there are four categories in three EOC situations. The average recognition rate and variance are shown in Table V. In the situation of SOC, the results of DeepEMD [28] and DeepEMD grid and DeepEMD sample in [27] are presented and it can be deduced that the initial DeepEMD methods are not suitable for SAR vehicle target recognition. Similarly, the improvement in randomly sampling and grid choosing based on DeepEMD cannot achieve a high recognition rate in few-shot SAR vehicle recognition, and they perform even worse than the initial DeepEMD without improvements. Even when the number of training samples climbs to 10 or even 25, there is no corresponding increase in the recognition rate. In ten-way 25-shot, the recognition rate is almost equal to ten-way ten-shot. This is mainly because the updated prototype representation in the SFC layer of DeepEMD cannot indicate the prototype of SAR vehicles in this category. Thus, when there is more labeled data because the number of training samples reaches 25 per class, only the variance decreases; the recognition rate does not increase.

2) *Experiments in EOCs*: This phenomenon also appears in the experiments in EOC1, EOC2/C, and EOC2/V. This approach uses the CNN to extract features and a relation network to measure the distance. However, the cascade representations of support images and query images cannot suitably reflect the distance between different categories of SAR vehicles, due to the sensitivity of azimuth angle. Therefore, the accuracy of relation networks in few-shot SAR vehicle recognition is poor in SOC and EOCs conditions. DN4 [38] and prototypical networks [37] show better performance than other approaches, except for DKTS-N. The recognition rate rises when the amount of labeled training data increases. This is because the DN4 used the NN classification module in the algorithm, which is suitable for SAR vehicles. Meanwhile, as the algorithm of prototypical networks, the mean of samples' feature works as the prototype to the category. Although the prototypical networks are more suitable than the SFC in DeepEMD, it is evident that when the number of training samples increases, the rise in accuracy is not as great as in our DKTS-N. From the results, it can be seen that even though there are only four categories of EOCs, they achieve a lower recognition rate than SOC.

After comparing the results from SOC and EOCs, it is obvious that the few-shot recognition tasks in EOCs are much more complicated than in SOC. This demonstrates that the changes in depression angle and configurations create differences in SAR images. Different configurations and versions in equipment contribute to the differences in component structure, which influences the scattering properties in SAR

TABLE V
RESULTS OF SOC AND EOCs AMONG FEW-SHOT LEARNING ALGORITHMS

SOC					
Algorithms	10-way 1-shot	10-way 2-shot	10-way 5-shot	10-way 10-shot	10-way 25-shot
DeepEMD [28]	36.19±0.46	43.49±0.44	53.14±0.40	59.64±0.39	59.71±0.31
DeepEMD grid [27]	35.89±0.43	41.15±0.41	52.24±0.37	56.04±0.31	57.89±0.24
DeepEMD sample [27]	35.47±0.44	42.39±0.42	50.34±0.39	52.36±0.28	55.02±0.22
DN4 [38]	33.25±0.49	44.15±0.45	53.48±0.41	64.88±0.34	79.28±0.22
Prototypical Network [37]	40.94±0.47	54.54±0.44	69.42±0.39	78.01±0.29	84.96±0.22
Relation Network [69]	36.19±0.46	43.49±0.44	53.14±0.40	59.64±0.39	59.71±0.31
Our DKTS-N	49.26±0.48	58.51±0.42	72.32±0.32	84.59±0.24	96.15±0.08
EOC1					
Algorithms	4-way 1-shot	4-way 2-shot	4-way 5-shot	4-way 10-shot	4-way 25-shot
DeepEMD [28]	56.81±0.99	62.8±0.78	65.16±0.61	67.58±0.49	70.22±0.35
DeepEMD grid [27]	55.95±0.43	57.46±0.41	63.81±0.37	65.72±0.31	68.85±0.24
DeepEMD sample [27]	49.65±0.44	54.00±0.42	58.19±0.39	60.34±0.28	62.51±0.22
DN4 [38]	46.59±0.83	51.41±0.69	58.11±0.49	62.15±0.43	65.14±0.37
Prototypical Network [37]	53.59±0.93	56.57±0.53	61.94±0.48	65.13±0.43	69.81±0.36
Relation Network [69]	43.21±1.02	46.93±0.81	54.97±0.56	38.62±0.49	44.42±0.43
Our DKTS-N	61.91±0.91	63.94±0.73	67.43±0.48	71.09±0.41	78.94±0.31
EOC2/C					
Algorithms	4-way 1-shot	4-way 2-shot	4-way 5-shot	4-way 10-shot	4-way 25-shot
DeepEMD [28]	38.39±0.86	45.65±0.75	54.53±0.60	62.13±0.50	63.71±0.36
DN4 [38]	46.13±0.69	51.21±0.62	58.14±0.54	63.08±0.51	69.66±0.46
Prototypical Network [37]	43.59±0.84	51.17±0.78	59.15±0.70	64.15±0.61	69.95±0.50
Relation Network [69]	42.13±0.90	48.24±0.82	53.12±0.71	36.28±0.59	39.81±0.42
Our DKTS-N	47.26±0.79	53.61±0.70	62.23±0.56	68.41±0.51	74.51±0.36
EOC2/V					
Algorithms	4-way 1-shot	4-way 2-shot	4-way 5-shot	4-way 10-shot	4-way 25-shot
DeepEMD [28]	40.92±0.76	49.12±0.65	58.43±0.51	67.64±0.42	67.03±0.21
DN4 [38]	47.00±0.72	52.21±0.61	58.87±0.55	63.93±0.52	70.64±0.47
Prototypical Network [37]	45.13±0.72	52.86±0.65	62.07±0.52	67.71±0.40	73.41±0.31
Relation Network [69]	40.24±0.91	46.32±0.82	54.22±0.68	35.13±0.52	33.81±0.46
Our DKTS-N	48.91±0.70	55.14±0.58	65.63±0.49	70.18±0.42	76.97±0.35

images. Because of this, recognition in few-shot situations is often problematic. Due to the randomly sampled images in the support set, the final results are influenced by the quality, the azimuth angle, and the shadow of images. As previously discussed, classification for four classes is more manageable than for ten classes, but our DKTS-N algorithm still experiences a nearly 20% drop from SOC to EOCs in 25-shot.

C. Ablation Experiments

In the ablation experiments, objective conditions were modified to control variables to show the effectiveness of our improvements in data augmentation and algorithm design. To make a fair comparison of the different improvements, most of the ablation experiments were conducted under SOC situations, except for researching the influence of depression angle (EOC1). Meanwhile, most of the experiments were ten-way one-shot, except for the comparison of SFC and the NN in time consumption and accuracy. The comparison involved a transformation from optical dataset or simulated SAR dataset, with and without phase data, with and without image stretching, the feature dimension, the improvement in measuring structural information in global and local features, and N -shot classification mode. Traditional classifiers are also

compared with our DKTS-N with the fixed features from the same CNNs. In addition, the estimated error of azimuth angle is considered in the ablation experiments.

1) *Comparison Between Configurations and Source Tasks:* Table VI shows that DKTS-N learns transfer information for SAR vehicle recognition much better in the SARSIM dataset than in the Mini-ImageNet dataset [33], which is a noted optical image dataset and sampled from ImageNet [10]. This also proves that the simulated SARSIM dataset is closer to MSTAR than Mini-ImageNet, meaning that SARSIM is a relatively ideal dataset for sampling source tasks to few-shot SAR vehicle recognition. Furthermore, SAR RGB images, which take the phase data into consideration, achieved a higher recognition rate than regular grayscale images in all the experiments. Generally, data augmentation with image stretching can increase the recognition rate between 1% and 4% than without image stretching. After trial and error, the settings of feature dimensions, 32–64–96–128, indicating the output filters of the convolutional block modules, provide relatively better results than others. In fact, SAR targets do not require very high dimensions as optical images due to their simpler image structure. Few-shot recognition methods for optical images tend to extract context information from the interested

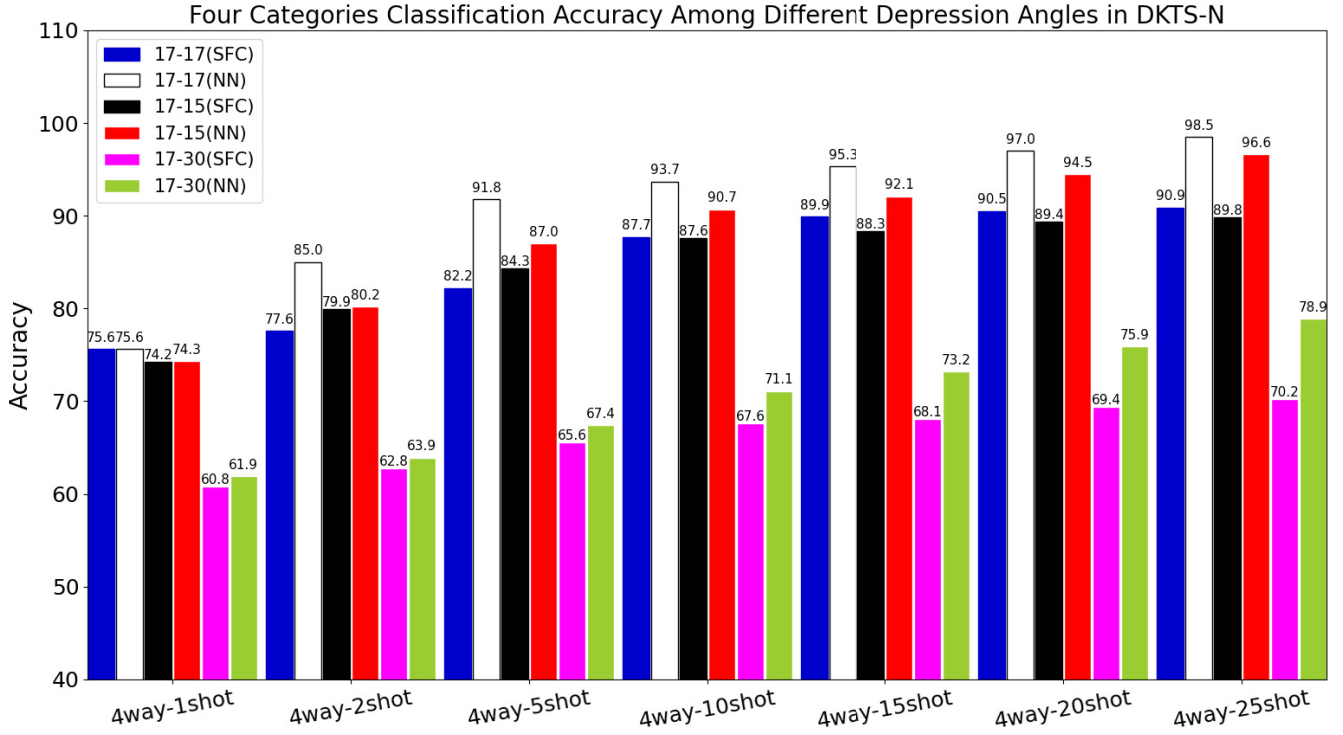


Fig. 5. Comparison of SFC and NN among different depression angles.

TABLE VI
INFLUENCE OF SOURCE TASKS AMONG SOC IN TEN-WAY ONE-SHOT

Source Task	Adjust	Form	Feature Dimension	Accuracy
Mini-ImageNet	✓	Grey	64-160-320-640	28.91±0.38
Mini-ImageNet	✗	Grey	64-160-320-640	29.128±0.27
Mini-ImageNet	✓	RGB	64-160-320-640	33.87±0.37
Mini-ImageNet	✗	RGB	64-160-320-640	39.83±0.39
SARSIM	✓	Grey	64-160-320-640	41.66±0.46
SARSIM	✗	Grey	64-160-320-640	28.19±0.48
SARSIM	✓	RGB	64-160-320-640	44.11±0.43
SARSIM	✗	RGB	64-160-320-640	42.04±0.50
SARSIM	✓	Grey	32-64-128-256	42.71±0.44
SARSIM	✗	Grey	32-64-128-256	38.92±0.51
SARSIM	✓	RGB	32-64-128-256	46.88±0.50
SARSIM	✗	RGB	32-64-128-256	43.03±0.49
SARSIM	✓	Grey	32-64-96-128	42.67±0.42
SARSIM	✗	Grey	32-64-96-128	38.93±0.49
SARSIM	✓	RGB	32-64-96-128	49.26±0.48
SARSIM	✗	RGB	32-64-96-128	44.22±0.44

targets, while DKTS-N aims to learn the structural information of global and local features in SAR targets.

2) *Effectiveness of NN in Accuracy and Time Consumption*: Table VII shows the different effects between SFC and NN, as well as time consumption. The multishot SAR vehicle recognition results of DKTS-N are also listed. First of all, the results prove that the recognition results of RGB SAR images, which contain both amplitude data and phase data, are better than grayscale SAR images. The bottom rows show the

results of DKTS-N in terms of accuracy and time consumption, based on stretched and RGB images. The two-, five-, and ten-shot recognition examples show that SFC takes more time and has lower recognition accuracy. The ten-shot recognition rates of the four settings are even lower than the five-shot ones in the SFC classifier. This means that the SFC classifier with updates is not suitable for SAR vehicle recognition. At the same time, the apparent increase can be seen in the NN classifier as the number of training samples increases. When it comes to ten-way ten-shot, the DKTS-N only needs 6.21 s per iteration (s/it), which means that each training or evaluation iteration takes 6.21 s, but for DeepEMD with SFC and NN, it takes 43.2 and 8.21 s/it, respectively. Due to our design in global and local features, the feature dimension is decreased, meaning that the calculation of EMD is simplified, and less time is needed. Overall, DKTS-N performs better in both accuracy and time consumption than DeepEMD.

3) *Influence of Depression Angle*: Although azimuth angles influence the few-shot recognition of SAR targets, depression angles have a greater effect. To measure the difference, four categories of 2S1, BRDM2, T72, and ZSU234 were fixed in the experiments. All the experiments shown in Fig. 5 are four-way with different shots. All the support images were sampled from sets at a 17° depression angle, whereas query images were selected from the rest of the images at a 17° depression angle, a 15° depression angle (SOC), and a 30° depression angle (EOC1). In Fig. 5, these are recorded as 17–17, 17–15, and 17–30. SFC and NN indicate the classifier module in DKTS-N. From the results, it can be seen that after the amount of training data increases to 15, it is difficult for SFC to increase the recognition rate as the number of training

TABLE VII
COMPARISON OF DKTS-N AND EMD AMONG SOC IN TEN-WAY N -SHOT

Adjust	Form	Settings	Algorithms	Classification	Time Consumption	Accuracy
✓	Grey	10-way 2-shot	DeepEMD	SFC / NN	14.01s/it / 1.94s/it	49.24±0.62 / 51.79±0.45
✗	Grey	10-way 2-shot	DeepEMD	SFC / NN	14.01s/it / 1.94s/it	44.12±0.52 / 46.50±0.41
✓	RGB	10-way 2-shot	DeepEMD	SFC / NN	14.01s/it / 1.94s/it	55.10±0.52 / 57.82±0.40
✗	RGB	10-way 2-shot	DeepEMD	SFC / NN	14.01s/it / 1.94s/it	54.02±0.49 / 56.51±0.42
✓	Grey	10-way 5-shot	DeepEMD	SFC / NN	19.8s/it / 5.87s/it	52.97±0.79 / 55.48±0.40
✗	Grey	10-way 5-shot	DeepEMD	SFC / NN	19.8s/it / 5.87s/it	49.95±0.71 / 54.27±0.38
✓	RGB	10-way 5-shot	DeepEMD	SFC / NN	19.8s/it / 5.87s/it	69.54±0.36 / 69.18±0.31
✗	RGB	10-way 5-shot	DeepEMD	SFC / NN	19.8s/it / 5.87s/it	64.44±0.35 / 66.24±0.32
✓	Grey	10-way 10-shot	DeepEMD	SFC / NN	43.2s/it / 8.21s/it	43.63±0.45 / 77.45±0.30
✗	Grey	10-way 10-shot	DeepEMD	SFC / NN	43.2s/it / 8.21s/it	52.13±0.51 / 66.04±0.29
✓	RGB	10-way 10-shot	DeepEMD	SFC / NN	43.2s/it / 8.21s/it	58.03±0.27 / 82.50±0.28
✗	RGB	10-way 10-shot	DeepEMD	SFC / NN	43.2s/it / 8.21s/it	56.49±0.63 / 80.15±0.27
✓	RGB	10-way 1-shot	DKTS-N	NN	1.15s/it	49.26±0.48
✓	RGB	10-way 2-shot	DKTS-N	NN	1.35s/it	58.51±0.42
✓	RGB	10-way 5-shot	DKTS-N	NN	3.81s/it	72.32±0.32
✓	RGB	10-way 10-shot	DKTS-N	NN	6.21s/it	84.59±0.24
✓	RGB	10-way 20-shot	DKTS-N	NN	9.12s/it	93.00±0.13
✓	RGB	10-way 25-shot	DKTS-N	NN	9.48s/it	96.15±0.11

TABLE VIII
COMPARISON EXPERIMENT OF CLASSICAL CLASSIFIERS

Algorithms	10-way 1-shot	10-way 2-shot	10-way 5-shot	10-way 10-shot	10-way 20-shot	10-way 25-shot
SVM [70]	38.75±0.45	50.32±0.41	67.49±0.34	77.99±0.27	85.99±0.20	87.85±0.18
LR [71]	41.96±0.35	52.82±0.37	69.06±0.32	79.52±0.24	86.98±0.19	88.84±0.18
DT [72]	18.54±0.47	26.02±0.44	40.70±0.43	50.31±0.38	55.67±0.33	57.37±0.32
GBC [73]	34.64±0.41	36.49±0.40	38.72±0.39	47.56±0.36	62.16±0.31	65.88±0.25
RF [74]	18.64±0.50	24.79±0.43	39.96±0.38	51.86±0.35	60.71±0.29	62.74±0.27
Our DKTS-N	49.26±0.48	58.51±0.42	72.32±0.32	84.59±0.24	93.00±0.13	96.15±0.08

images increases. Because in the one-shot classification, there is no difference between NN and SFC, the results are close to each other. On the whole, the query images sampled from 17 (for the sake of simplicity, recorded as 17–17) come the closest to the support images in the depression angle, and thus, the recognition rate is the highest in all the experiments. When it comes to the 17–15, the depression angle changes a little, and thus, small variations will occur in the SAR images. This contributes to the small decrease in the recognition rate, as the dotted lines (17–17) are a bit higher than the solid lines (17–15) in both SFC and NN. However, when the depression angle changes only slightly, from 15° to 30°, the scattering structures and positions change in the SAR images. This can be a critical obstacle in few-shot SAR vehicle recognition. As shown in Fig. 5, the results of 17–30 are much lower than 17–17 and 17–15, which also proves that the few-shot SAR vehicle recognition is highly sensitive to the depression angles.

4) *Comparison With Classical Classifiers:* The proposed DKTS-N algorithm aims to solve the few-shot problem in SAR vehicle target recognition. Comparative experiments with support vector machine (SVM) [70], logistic regression (LR) [71], decision tree (DT) [72], gradient boosting classifier (GBC) [73], and random forest (RF) [74] algorithms were conducted under the premise that only the classifiers are

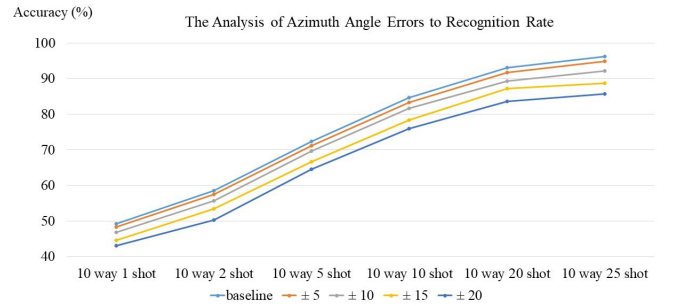


Fig. 6. Line chart of different azimuth angle errors to the recognition rate.

changed and the extracted features are the same as DKTS-N. In Table VIII, all the compared algorithms are lower in performance than DKTS-N. The running times of some of these classical algorithms are slightly influenced by the shot numbers, but they are acutely affected by the dimension of features. The running times of LR and GBC are all about ten times longer than that of DKTS-N, while SVM consumes half time of DKTS-N and has relatively good performance among traditional classifiers as well.

5) *Estimated Errors of Azimuth Angle:* The results of experiments under the different azimuth angle errors are shown in Fig. 6. For the five groups of experiments, all the

configurations were the same except for the range of random azimuth angle error. ± 5 indicates that the estimation error of rotation angle is from -5° to 5° . It is easy to see that when the estimation error is less than $\pm 5^\circ$, the recognition rates experience a nearly 1% decrease compared to the ideal situation. When the angle error reaches $\pm 20^\circ$, the recognition rates drop about 10%. The baseline in the figure means that the azimuth angle estimation error is nearly zero.

VI. CONCLUSION

In this study, the DKTS-N algorithm is proposed as an effective solution to the problem of few-shot SAR vehicle recognition. SAR domain knowledge, based on complex information and azimuth angle, is taken into consideration in the data augmentation process. The augmentation framework extends grayscale SAR images to three-channel RGB data and rotates them through the method of azimuth angle normalization, which is an effective and pioneering work in few-shot SAR vehicle recognition. Furthermore, a two-stream deep network is used to extract the global and local features of SAR vehicle targets. To measure the structural information for the above-mentioned features, EMD is improved to fit the situation in this study. After performing a number of experiments, designed to verify the potential of DKTS-N in K -shot target classification, the NN classifier was confirmed to perform better than the SFC classifier in SAR vehicle recognition. Experiments used the simulated dataset Sarsim as the source of source tasks and the MSTAR dataset (SOC, EOC1, EOC2/C, and EOC2/V) as the source of target tasks, which approximated the real-world applications and was easily repeatable. Comparative experiments proved that DKTS-N performed better than the DeepEMD and other few-shot or regular classification methods.

One of the problems related to few-shot SAR vehicle recognition is that algorithms are still sensitive to the azimuth angle and depression angle, which is confirmed by the results of experiments. Our proposed method is more suitable for vehicles than ships and planes. SAR planes are more sensitive to the azimuth angle than vehicles, and the global and local features of planes with different azimuth angles have little information in common. Furthermore, SAR ship targets and plane targets have a wide range in target size. As a result, it is difficult to compare all this information through the proposed method in this study. In future work, we will try to find other useful features of SAR vehicles, such as the imaging time sequences and scattering characteristics of equipment in these vehicles, in order to further improve the performance of algorithms.

REFERENCES

- [1] O. Kechagias-Stamatis and N. Aouf, "Automatic target recognition on synthetic aperture radar imagery: A survey," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 3, pp. 56–81, Mar. 2021.
- [2] 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 Trans. Geosci. Remote. Sens. Lett.*, vol. 14, no. 7, pp. 1091–1095, Jul. 2017.
- [3] T. Zhuangzhuang, Z. Ronghui, H. Jiemin, and Z. Jun, "SAR ATR based on convolutional neural network," *J. Radars*, vol. 5, no. 3, pp. 320–325, Mar. 2016.
- [4] G. Weiwei, Z. Zenghui, Y. Wenxian, and S. Xiaohua, "Perspective on explainable SAR target recognition," *J. Radars*, vol. 9, no. 3, pp. 462–476, 2020.
- [5] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class mstar data set," *Appl. Sci.*, vol. 6, no. 1, pp. 26–35, 2016.
- [6] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, pp. 192–201, 2017.
- [7] 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, Aug. 2016.
- [8] G. Dong, G. Kuang, N. Wang, and W. Wang, "Classification via sparse representation of steerable wavelet frames on Grassmann manifold: Application to target recognition in SAR image," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2892–2904, Jun. 2017.
- [9] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [11] L. Liu *et al.*, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 1–58, 2018.
- [12] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.
- [13] R. Chakraborty, J. Wang, and S. X. Yu, "SurReal: Fréchet mean and distance transform for complex-valued deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–9.
- [14] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T. S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Trans. Geoscience Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.
- [15] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.
- [16] D. Lu, L. Cao, and H. Liu, "Few-shot learning neural network for SAR target recognition," in *Proc. 6th Asia-Pacific Conf. Synth. Aperture Radar (APSAR)*, Nov. 2019, pp. 1–4.
- [17] W. D. E. A. Yikui Zhai, "MFFA-SARNET: Deep transferred multi-level feature fusion attention network with dual optimized loss for small-sample SAR ATR," *Remote Sens.*, vol. 12, no. 9, pp. 1385–1396, 2020.
- [18] J. Tang, F. Zhang, Y. Zhou, Q. Yin, and W. Hu, "A fast inference networks for SAR target few-shot learning based on improved Siamese networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 1212–1215.
- [19] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "SAR image classification using few-shot cross-domain transfer learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–9.
- [20] L. Zeng, D. Zhou, X. Li, and Z. Kun, "Novel SAR target detection algorithm using free training," *J. Radars*, vol. 6, no. 2, pp. 177–185, 2017.
- [21] Z. Huang, Z. Pan, and B. Lei, "What, where, and how to transfer in SAR target recognition based on deep CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2324–2336, Apr. 2020.
- [22] K. El-Darymli, P. McGuire, D. Power, and C. Moloney, "Rethinking the phase in single-channel SAR imagery," in *Proc. Int. Radar Symp.*, Jun. 2013, pp. 429–436.
- [23] P. Zhao, K. Liu, H. Zou, and X. Zhen, "Multi-stream convolutional neural network for SAR automatic target recognition," *Remote Sens.*, vol. 10, no. 9, pp. 1473–1484, 2018.
- [24] H. Zou, L. Yun, and H. Wen, "Research on multi-aspect SAR images target recognition using deep learning," *J. Signal Process.*, vol. 34, no. 5, pp. 512–522, 2018.
- [25] R. Xue, X. Bai, and F. Zhou, "Spatial-temporal ensemble convolution for sequence SAR target classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1250–1262, Feb. 2020.
- [26] F. Zhang, Z. Fu, Y. Zhou, W. Hu, and W. Hong, "Multi-aspect SAR target recognition based on space-fixed and space-varying scattering feature joint learning," *Remote Sens. Lett.*, vol. 10, no. 10, pp. 998–1007, 2019.
- [27] C. Zhang, Y. Cai, G. Lin, and C. Shen, "DeepEMD: Differentiable Earth mover's distance for few-shot learning," 2020, *arXiv:2003.06777*. [Online]. Available: <http://arxiv.org/abs/2003.06777>

- [28] C. Zhang, Y. Cai, G. Lin, and C. Shen, "DeepEMD: Few-shot image classification with differentiable Earth mover's distance and structured classifiers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12203–12213.
- [29] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.
- [30] E. Keydel, S. Lee, and J. Moore, "MSTAR extended operating conditions: A tutorial," *Proc. SPIE*, vol. 2757, pp. 228–242, Jun. 1996.
- [31] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1484–1488, Sep. 2017.
- [32] 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.
- [33] A. Srinivasan, A. Bharadwaj, M. Sathyan, and S. Natarajan, "Optimization of image embeddings for few shot learning," in *Proc. 10th Int. Conf. Pattern Recognit. Appl. Methods*, 2021, pp. 1–6.
- [34] B. M. Lake, R. Salakhutdinov, J. Gross, and J. Tenenbaum, "One shot learning of simple visual concepts," in *Proc. Annu. Meeting Cogn. Sci. Soc.*, vol. 33, 2011, pp. 2568–2573.
- [35] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [36] O. Vinyals, C. Blundell, T. P. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Neural Inf. Process. Syst.*, 2016, pp. 3637–3645.
- [37] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. Neural Inf. Process. Syst.*, 2017, pp. 4077–4087.
- [38] W. Li, L. Wang, J. Xu, J. Huo, and J. Luo, "Revisiting local descriptor based image-to-class measure for few-shot learning," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7260–7268.
- [39] X. Yang, X. Nan, and B. Song, "D2N4: A discriminative deep nearest neighbor neural network for few-shot space target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3667–3676, May 2020.
- [40] L. Wang, X. Bai, and F. Zhou, "Few-shot SAR ATR based on conv-BiLSTM prototypical networks," in *Proc. 6th Asia-Pacific Conf. Synth. Aperture Radar (APSAR)*, Nov. 2019, pp. 1–5.
- [41] K. Fu, T. Zhang, Y. Zhang, Z. Wang, and X. Sun, "Few-shot SAR target classification via metalearning," *IEEE Trans. Geosci. Remote Sens.*, early access, Feb. 22, 2021, doi: [10.1109/TGRS.2021.3058249](https://doi.org/10.1109/TGRS.2021.3058249).
- [42] L. Wang, X. Bai, C. Gong, and F. Zhou, "Hybrid inference network for few-shot SAR automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 22, 2021, doi: [10.1109/TGRS.2021.3051024](https://doi.org/10.1109/TGRS.2021.3051024).
- [43] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, vol. 2, 2015, pp. 1–5.
- [44] Z. Zhang, H. Wang, F. Xu, and Y. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.
- [45] X. Feng, W. Haipeng, and J. Yaqiu, "Deep learning as applied in SAR target recognition and terrain classification," *J. Radars*, vol. 6, no. 2, pp. 136–148, 2017.
- [46] K. El-Darymli, C. Moloney, E. Gill, P. McGuire, and D. Power, "On circularity/noncircularity in single-channel synthetic aperture radar imagery," in *Proc. Oceans-St. John's*, Sep. 2015, pp. 1–4.
- [47] X. Leng, K. Ji, S. Zhou, and X. Xing, "Ship detection based on complex signal kurtosis in single-channel SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6447–6461, Sep. 2019.
- [48] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in SAR images by exploiting the azimuth sensitivity," *Remote Sens. Lett.*, vol. 8, no. 9, pp. 821–830, 2016.
- [49] S. Schuster, P. Vernaza, W. Choi, and M. Chandraker, "Deep network flow for multi-object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6951–6960.
- [50] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 957–966.
- [51] Q. Zhao, Z. Yang, and H. Tao, "Differential Earth mover's distance with its applications to visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 274–287, Feb. 2010.
- [52] C. Wang and S. C. Chan, "A new hand gesture recognition algorithm based on joint color-depth superpixel Earth mover's distance," in *Proc. 4th Int. Workshop Cognit. Inf. Process. (CIP)*, May 2014, pp. 1–6.
- [53] P. Li, "Tensor-SIFT based Earth mover's distance for contour tracking," *J. Math. Imag. Vis.*, vol. 46, no. 1, pp. 44–65, May 2013.
- [54] B. Hou, N. Li, S. Wang, and X. Zhang, "SAR image segmentation based on random projection and signature frame," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 3726–3729.
- [55] W. Xu, G. Zhang, and Y. Duan, "Farmland detection in synthetic aperture radar images with texture signature," *J. Appl. Remote Sens.*, vol. 8, no. 1, Mar. 2014, Art. no. 084997.
- [56] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions Solution Mappings*. New York, NY, USA: Springer, 2009.
- [57] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem: History Theory Application*. Basel, Switzerland: Birkhauser, 2003.
- [58] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 79–91, Jan. 1997.
- [59] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3334–3347, Jul. 2017.
- [60] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Data augmentation by multilevel reconstruction using attributed scattering center for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 979–983, Jun. 2017.
- [61] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [62] F. L. Hitchcock, "The distribution of a product from several sources to numerous localities," *J. Math. Phys.*, vol. 20, nos. 1–4, pp. 224–230, Apr. 1941.
- [63] *Sensor Data Management System Website, MSTAR Database*. [Online]. Available: <https://www.sdms.af.mil/index.php?collection=mstar>
- [64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [66] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on esenet," *Remote Sens.*, vol. 11, no. 11, pp. 1316–1325, 2019.
- [67] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.
- [68] H. Y. Wano, P. Zongxu, and M. Wen, "SAR target recognition based on Siamese CNN with small scale dataset," *Radar Sci. Technol.*, vol. 17, no. 6, pp. 603–609, 2019.
- [69] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.
- [70] A. Ukil, "Support vector machine," *Comput. Sci.*, vol. 1, no. 4, pp. 1–28, 2002.
- [71] P. D. Allison, *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC, USA: SAS, 1999.
- [72] D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man Cybern.*, vol. 21, no. 3, pp. 660–674, May 1991.
- [73] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4626–4636, Apr. 2009.
- [74] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 23, no. 23, pp. 18–22, 2002.



Linbin Zhang received the B.S. degree in information engineering from the National University of Defense Technology, Changsha, China, in 2018, where he is currently pursuing the Ph.D. degree with the State Key Laboratory of Complex Electromagnetic Environment Effects.

His research interests include object detection, image classification, few-shot learning, and machine learning and its applications to remote sensing images.



Xiangguang Leng (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University (WHU), Wuhan, China, in 2013, and the M.S. degree (Hons.) in photogrammetry and remote sensing and the Ph.D. degree in electronic science and technology from the National University of Defense Technology (NUDT), Changsha, China, in 2015 and 2019, respectively.

He is currently a Lecturer at the College of Electronic Science and Technology, NUDT. He has published more than 30 articles. His research interests mainly include remote sensing information processing, synthetic aperture radar (SAR) image interpretation, and machine learning.

Dr. Leng was a recipient of the Excellent Paper Award of the 2016 CIE International Conference on Radar, the Excellent Paper Award of the 2018 IET International Radar Conference, and the Excellent Paper Award of the 5th China High Resolution Earth Observation Conference. He is a reviewer of several international journals.



Sijia Feng received the B.S. and M.S. degrees in information engineering from the National University of Defense Technology, Changsha, China, in 2016 and 2019, respectively, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Complex Electromagnetic Environment Effects.

Her research interests include synthetic aperture radar (SAR) image interpretation, feature extraction, and machine learning.



Xiaojie Ma received the B.S. and M.S. degrees from Yanshan University, Qinhuangdao, China, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Complex Electromagnetic Environment Effects, National University of Defense Technology, Changsha, China.

His research interests include synthetic aperture radar (SAR) automatic target recognition (ATR) and machine learning.



Kefeng Ji (Member, IEEE) received the B.S. degree in aerospace engineering from Northwestern Polytechnical University (NWPU), Xi'an, China, in 1996, and the M.S. and Ph.D. degrees in information and telecommunication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 1999 and 2003, respectively.

In 2003, he joined the College of Electronic Science and Technology, NUDT, where he is currently a Professor. He has authored or coauthored over 80 articles. His research interests include signal processing, machine learning, pattern recognition, remote sensing information processing, synthetic aperture radar (SAR) image interpretation, target detection, recognition, feature extraction, and marine surveillance.

Dr. Ji is a Reviewer of several international journals and conferences, such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE ACCESS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *International Journal of Remote Sensing*, *Remote Sensing Letters*, *European Journal of Remote Sensing*, and *Remote Sensing*.



Gangyao Kuang (Senior Member, IEEE) received the B.S. and M.S. degrees in geophysics from the Central South University of Technology, Changsha, China, in 1988 and 1991, respectively, and the Ph.D. degree in communication and information from the National University of Defense Technology, Changsha, in 1995.

He is currently a Professor at the School of Electronic Science, National University of Defense Technology. His research interests include remote sensing, synthetic aperture radar (SAR) image processing, change detection, SAR ground moving target indication, and classification with polarimetric SAR images.



Li Liu (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2012.

She is currently a Professor with the College of System Engineering. During her Ph.D. study, she spent more than two years as a Visiting Student at the University of Waterloo, Waterloo, ON, Canada, from 2008 to 2010. From 2015 to 2016, she spent ten months visiting the Multimedia Laboratory at The Chinese University of Hong Kong, Hong Kong. From December 2016 to November 2018, she worked as a Senior Researcher at the Machine Vision Group, University of Oulu, Oulu, Finland. Her articles have currently over 4700 citations in Google Scholar. Her research interests include computer vision, pattern recognition, and machine learning.

Dr. Liu served as the Guest Editor for special issues in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI) and *International Journal of Computer Vision* (IJCV) and serves as an Associate Editor for *Pattern Recognition* and *Pattern Recognition Letters*.