Few-Shot SAR Target Classification Combining Both Spatial and Frequency Information

Haorun Li , Tianyu Wang, and Shaowei Wang
School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
Email: dg21230079@smail.nju.edu.cn, {tianyu.alex.wang, wangsw}@nju.edu.cn

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001168

***Abstract*—The automatic recognition of synthetic aperture radar (SAR) targets has been extensively studied in recent years. Specifically, due to the high cost of SAR image acquisition and the low occurrence probability of high-value targets, it is of great importance to identify SAR targets with only a few available images, which is referred to as few-shot SAR target classification. However, most existing solutions straightly adopt meta-learning and transfer learning methods that are originally designed for optical images, which do not consider the unique frequency information of SAR images. In this paper, we propose a novel hybrid classification network that combines both spatial and frequency information for few-shot SAR target classification. Specifically, we first train the proposed network in a source dataset, which contains a large number of related SAR images without the targets of interest. Then, the pre-trained network is fine-tuned on the target dataset consisting of only a few SAR images of interest. Compared with the conventional method based on convolutional neural networks and model-agnostic meta-learning, the proposed method can achieve superior top-1 accuracy in various settings.**

***Index Terms*—Big data, few-shot learning, frequency information, SAR target classification, transfer learning.**

I. Introduction

Synthetic aperture radar (SAR) refers to a unique implementation of a radar system, which utilizes the movement of the radar platform and dedicated signal processing techniques to generate high-resolution images. Compared with the optical imaging system, the SAR imaging system has the capability of all-weather and all-day surveillance. Therefore, SAR plays an important role in many applications, including marine observation, geological exploration, and terrain mapping [1]. On the one hand, due to the special imaging mechanism of active radar systems, SAR images can provide unique information that can not be reflected in optical images, e.g., the backscattering patterns. On the other hand, due to the extra information provided by SAR images, SAR target recognition is considered to be more complicated and more challenging. Therefore, the SAR target classification has attracted an intensive attention in the literature [2]–[5].

Traditional methods for the SAR target classification can be roughly classified into three categories, i.e., template matching, target modeling, and kernel-based classifiers [2]. However, these methods heavily rely on hand-designed features, which results in poor generalization performance, low efficiency, and high complexity. With the advent of big data processing technologies in recent years, deep neural networks have been widely used in many research domains including SAR target classification [6]–[8]. Specifically, deep neural networks have shown technical advantages in learning hierarchical features and achieve great success in SAR target classification. In [3], a deep convolutional neural network is adopted for the high-resolution polarimetric SAR scene classification. In [4], the author presents a combination of convolutional neural networks and support vector machines for the single-polarized SAR target classification, which increases the generalization capability and keeps low inference time.

Although recent studies have shown deep neural networks have great advancements in the SAR target classification, these networks require a large number of SAR images for training. However, in practical applications, it can be difficult to obtain sufficient SAR images with fine annotation [5]. The reasons include: 1) The acquisition of SAR images can be expensive as SAR imaging devices are usually carried on aircraft or satellites. 2) The labeling of SAR images is a laborious and time-consuming task as the interpretation of SAR images is more difficult than that of optical images. 3) High-value observation targets are rarely detected. Therefore, it is important to identify SAR targets with only a few images of interest, which is referred to as few-shot SAR target classification.

In the traditional few-shot classification problem of optical images, there are two major methods, i.e., transfer learning and meta-learning. Transfer learning aims to resolve the target domain problem by leveraging the prior knowledge contained in the source domain [9]. According to the type of prior knowledge, the transfer learning approaches can be classified into four categories, i.e., relational-knowledge-transfer, instance-transfer, parameter-transfer, and feature-representation-transfer [10]. Relational-knowledge-transfer takes the relationships learned in the source domain as the prior knowledge. Instance-transfer focuses on reusing data in the source domain. Parameter-transfer encodes the prior knowledge into the network parameters and reuse such parameters in the target domain. Feature-representation-transfer aims to learn a good feature representation from the source domain, which can also be

This work was supported in part by the National Natural Science Foundation of China under Grants 61931023 and U1936202.

applied for the target domain problem.

Meta-learning aims to accumulate experience via learning numerous tasks in the source domain. The meta-learning techniques can be classified into three categories, i.e., metric-based, model-based, and optimization-based methods [11]. The metric-based methods learn a set of embedding functions such that data samples can be projected to a special space where samples with different labels can be easily separated [12]. The optimization-based methods learn to search for an optimal parameter initialization of a base-learner such that it can be fine-tuned effectively by using the few samples in the target domain [13]. The model-based methods aim to fast update parameters by a small number of labeled samples through specially designed networks [14].

Most of the existing studies on the few-shot SAR target classification focus on applying the existing few-shot learning methods in optical image classification. In [15], the authors propose an end-to-end few-shot SAR classification method to recognize SAR targets with only a few images. In [16], a conventional method based on convolutional neural networks and model-agnostic meta-learning is proposed to solve the few-shot SAR target classification problem. In [17], the authors replace the relation module in the relation network with a graph neural network to help improve the performance of few-shot SAR target classification. In these studies, only the spatial information of SAR images is utilized and the unique frequency information is not considered. However, frequency information can provide extra features of SAR targets, e.g., backscattering patterns and physical properties. Thus, some recent studies propose to involve the frequency information in SAR target classification. In [18], the spectral analysis of complex-valued SAR images is introduced for the target detection. In [19], the multiple-sublook decomposition method based on time-frequency analysis is proposed to emphasize the target characterization of complex-valued SAR images. In [20], a deep learning framework named Deep SAR-Net is proposed to make full use of complex-valued SAR data.

In this paper, we introduce a novel hybrid network combining both spatial and frequency information (HNCSF) to address the few-shot SAR target classification problem. We first design two different feature extraction networks, i.e., a residual network and a convolutional auto-encoder to obtain features from spatial and frequency information, respectively. Then, the spatial and frequency features are fused and sent into another residual network to output the classification results. Specially, we introduce the transfer learning technology to help train our proposed HNCSF. Experimental results show that our proposed HNCSF can effectively utilize both the spatial and frequency information and achieve high top-1 accuracy in various few-shot settings.

II. Data Pre-processing

A. Source and Target Datasets

We denote a 2-D complex-valued SAR image of size $N_{cx} \times N_{cy}$ as $\boldsymbol{C}(x,y) = \boldsymbol{A}(x,y) + j\boldsymbol{B}(x,y)$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are both real matrices, x and y represent the spatial positions of each pixel in range and azimuth directions, respectively. The label of $\boldsymbol{C}(x,y)$ is a one-hot vector and denoted as $\boldsymbol{L}(\boldsymbol{C}(x,y))$. In the few-shot SAR target classification, we need to clarify two datasets, i.e., the target dataset D_t and the source dataset D_s.

The target dataset D_t consists of the few available images of SAR targets that need to be identified. For a target dataset with N_t SAR target categories and K_t available images per target category, we refer to the corresponding few-shot problem as an N_t-way K_t-shot problem. Here, we construct the target dataset by using the popular S1 dataset [20]. Specifically, we choose three target categories from S1, i.e., water, container, and skyscraper, and randomly select 1, 5, and 10 images to construct a 3-way 1-shot problem, a 3-way 5-shot problem, and a 3-way 10-shot problem, respectively.

The source dataset D_s consists of SAR targets with a large number of available images. The SAR target categories in the source dataset are strictly different from the SAR target categories in the target dataset, while these two datasets should be highly related such that the source dataset can help improve the performance of the network on the target dataset. Here, we construct the source dataset by using all the images of the rest SAR target categories ($N_s = 5$) in the S1 dataset, i.e., agriculture, forest, industrial building, residential, and storage tank.

B. Spatial and Frequency Information

Given a 2-D complex-valued SAR image $\boldsymbol{C}(x,y)$ of size $N_{cx} \times N_{cy}$, the spatial information has the same dimension $[N_{cx}, N_{cy}]$ and is given by [17]

$$\boldsymbol{\mathcal{P}_s}(x,y) = \sqrt{\boldsymbol{A}(x,y)^2 + \boldsymbol{B}(x,y)^2}. \tag{1}$$

The spatial information $\boldsymbol{\mathcal{P}_s}(x,y)$ reveals the backscatter intensity, which represents the roughness of the target surface.

The frequency information is a 4-D representation of the 2-D SAR image $\boldsymbol{C}(x,y)$, which is given by [19]

$$\begin{aligned} &\boldsymbol{\mathcal{P}_f}(x,y,f_r,f_{az}) \\ &= \left|\mathrm{FFT}^{-1}[\boldsymbol{w}(f_r,f_{az}) \cdot \mathrm{FFT}(\boldsymbol{C})](x,y)\right| \end{aligned} \tag{2}$$

where FFT and FFT^{-1} represent the 2-D Fourier transform and inverse Fourier transform, respectively, (f_r, f_{az}) are the frequency coordinates in the $\mathrm{FFT}(\boldsymbol{C})$, and $\boldsymbol{w}(f_r,f_{az})$ is a bandpass filter centered at the frequencies (f_r,f_{az}) with the bandwidths of b_r and b_{az} in range and azimuth, respectively. The frequency information $\boldsymbol{\mathcal{P}_f}(x,y,f_r,f_{az})$ is closely related to the physical properties and backscattering patterns.

III. Architecture of HNCSF

The network structure of our proposed HNCSF is shown in Fig. 1, which consists of four components: the spatial feature extraction network $\mathcal{M}_s$, the frequency feature extraction network $\mathcal{M}_f$, the feature fusion network $\mathcal{V}$, and the SAR target classification network $\mathcal{G}$. As we can see, $\mathcal{M}_s$ takes the spatial information $\boldsymbol{\mathcal{P}_s}(x,y)$ as its input and outputs the spatial features $\boldsymbol{\Psi_s}$. $\mathcal{M}_f$ takes the frequency information

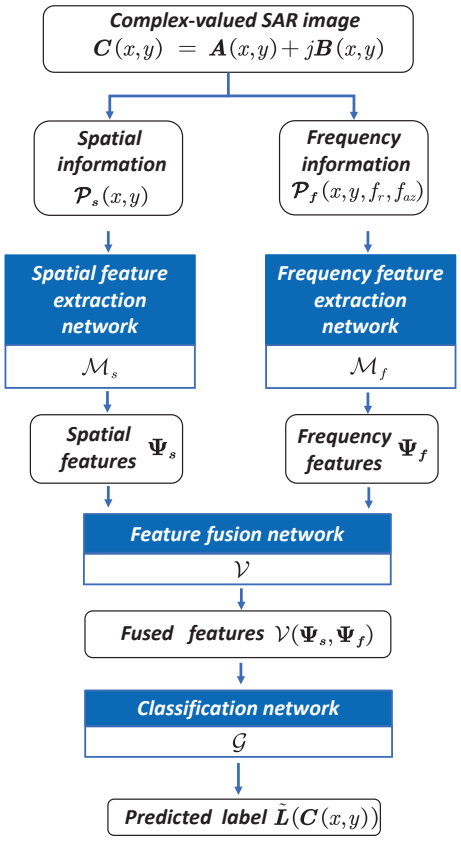

Fig. 1. The overall architecture of the proposed HNCSF.

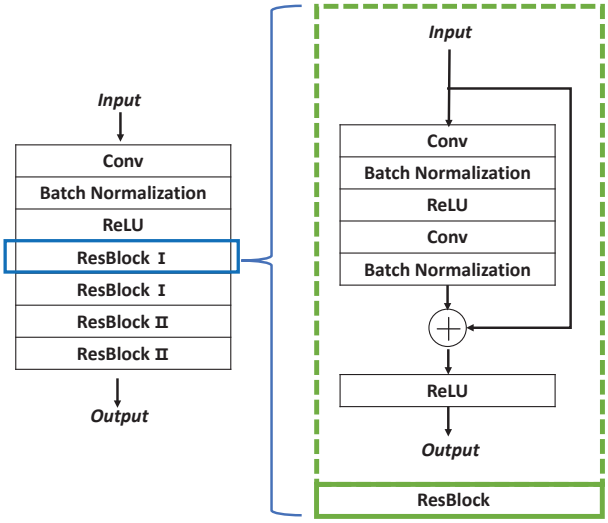

Fig. 2. The architecture of the spatial feature extraction network $\mathcal{M}_s$.

$\mathcal{P}_f(x, y, f_r, f_{az})$ as its input and outputs the frequency features Ψ_f. Then, Ψ_s and Ψ_f are sent into the feature fusion network $\mathcal{V}$ to generate the fused features $\mathcal{V}(\Psi_s, \Psi_f)$. At last, $\mathcal{G}$ takes $\mathcal{V}(\Psi_s, \Psi_f)$ as its input and outputs the predicted label $\tilde{L}(C(x, y))$ of the original input SAR image $C(x, y)$.

A. Spatial Feature Extraction Network

The spatial feature extraction network $\mathcal{M}_s$ mainly follows the conventional residual network architecture [21]. As shown in Fig. 2, $\mathcal{M}_s$ contains a convolutional layer followed by a batch normalization layer and a ReLU activation layer, as well as four residual blocks (ResBlock). Each ResBlock consists of an identity shortcut connection and two convolutional layers followed by batch normalization layers and ReLU activation layers. The first and the third ResBlocks are identical to the second and fourth ResBlocks, respectively.

The convolutional layers perform a convolutional operation of the input images by using a set of kernel filters to extract the low-dimensional features. Compared with the fully connected layers, the number of parameters to be learned of the conventional layers is reduced significantly. In the batch normalization layers, each input channel is subtracted by the mean of the batch and then divided by its standard deviation. The batch normalization layers not only accelerate the convergence of the network but also play an important role in solving the vanishing gradient problem. The ReLU activation layers are applied to increase the nonlinear properties by replacing all negative input values with zeros. The identity shortcut connection in the ResBlock allows the ReLU layer to directly utilize the features from the original input, which helps to resolve the predominant problem of vanishing gradient.

B. Frequency Feature Extraction Network

As shown in Fig. 3, the frequency feature extraction network $\mathcal{M}_f$ is a conventional convolutional neural network, which is made up of four different convolutional layers, each followed by a batch normalization layer and a ReLU activation layer. Besides, the max-pooling layers are attached after the first and third ReLU activation layers.

The convolutional layers, the batch normalization layers, and the ReLU activation layers have been explained in Section III-A. The max-pooling layers perform down-sampling by choosing the maximum value of the elements in the input feature maps. The main purpose of the max-pooling layers is to maintain the key features while reducing the number of parameters to be learned.

C. Feature Fusion Network

The feature fusion network $\mathcal{V}$ fuses Ψ_s and Ψ_f to generate $\mathcal{V}(\Psi_s, \Psi_f)$ by directly concatenating Ψ_s and Ψ_f together such that both the spatial and frequency information are considered in the network. The hyperparameters of $\mathcal{M}_s$ and $\mathcal{M}_f$ (e.g., kernel size and stride) should be specifically chosen to make sure that Ψ_s and Ψ_f have compatible sizes, i.e., the lengths of the dimensions match except for the concatenated dimension.

D. SAR Target Classification Network

As shown in Fig. 4, the SAR target classification network $\mathcal{G}$ is a specifically designed residual network, which consists of two identical residual bottleneck blocks (ResBottleneckBlock),

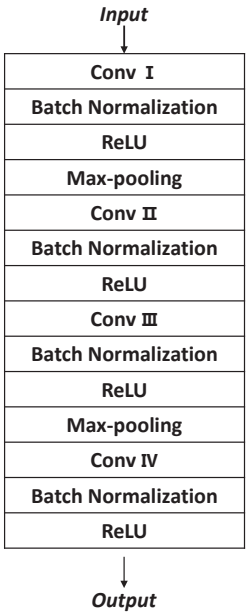

Fig. 3. The architecture of the frequency feature extraction network $\mathcal{M}_f$.

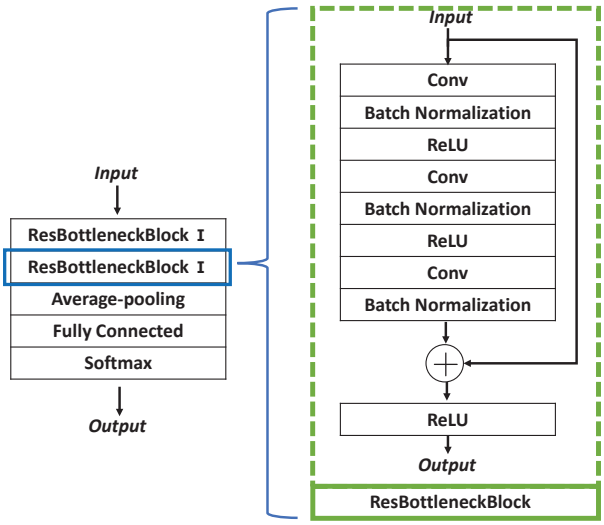

Fig. 4. The architecture of the SAR target classification network $\mathcal{G}$.

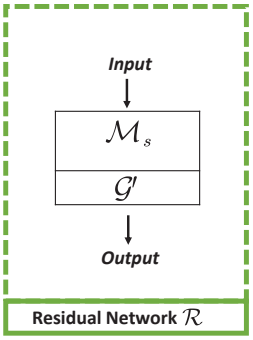

Fig. 5. The architecture of the residual network $\mathcal{R}$.

followed by an adaptive average-pooling layer, a fully connected layer, and a softmax layer. Each ResBottleneckBlock consists of an identity shortcut connection and three convolutional layers followed by batch normalization layers and ReLU activation layers.

The average-pooling layers calculate the average value of the elements in the input feature maps to achieve the same purpose as the max-pooling layers. In the fully connected layers, all the inputs are connected to every node of the next layer. The fully connected layers can map the learned features to the label space. The softmax layers are attached after the fully connected layers to output the probability distributions of the classification results. The original input image is classified into the class corresponding to the highest probability. The identity shortcut connection of the ResBottleneckBlock has the same effect as that of the ResBlock.

IV. Training Process Based on Transfer Learning

It is impractical to directly train the proposed HNCSF with such few available images in the target dataset. Hence, we adopt the transfer learning technology, which utilizes the inherent similarity between the SAR targets in the source and target datasets to help train the proposed HNCSF. Specifically, the training process can be divided into two stages, i.e., the pre-training stage on the source dataset and the fine-tuning stage on the target dataset. These two stages will be explained separately in the following paragraphs.

A. Pre-training

The pre-training stage aims to make full use of the prior knowledge by training the HNCSF on the source dataset. As shown in Fig. 5, we construct a residual network $\mathcal{R}$ by stacking the spatial feature extraction network $\mathcal{M}_s$ and the classification network $\mathcal{G}'$. $\mathcal{G}'$ has a similar structure to $\mathcal{G}$ except that the number of input channels is adapted to match the output dimension of $\mathcal{M}_s$.

$\mathcal{R}$ takes the spatial information $\boldsymbol{\mathcal{P}_s}(x, y)$ as its input and outputs the predicted label of the SAR image $\boldsymbol{C}$. The cross-entropy loss function is applied, which is given by

$$Loss_{\mathcal{R}} = -\frac{1}{|D_s|}\sum_{j=1}^{|D_s|}\sum_{i=1}^{N_s} L_i(\boldsymbol{C_j})\log_2(\tilde{L}_{\mathcal{R},i}(\boldsymbol{C_j})) \quad (3)$$

where $|D_s|$ represents the total number of SAR images of the source dataset, $L_i(\boldsymbol{C_j})$ represents the i-th element of the label vector $\boldsymbol{L}(\boldsymbol{C_j})$, and $\tilde{L}_{\mathcal{R},i}(\boldsymbol{C_j})$ represents the i-th element of the output of $\mathcal{R}$. Given the loss function $Loss_{\mathcal{R}}$, $\mathcal{R}$ is trained by using the stochastic gradient descent algorithm.

As shown in Fig. 6, we construct a convolutional auto-encoder $\mathcal{S}$, in which the encoder is the frequency feature extraction network $\mathcal{M}_f$ and the decoder $\tilde{\mathcal{M}}_f$ is symmetric with the encoder $\mathcal{M}_f$. $\mathcal{M}_f$ takes the frequency information $\boldsymbol{\mathcal{P}_f^j}$ as its input and outputs frequency features, which are

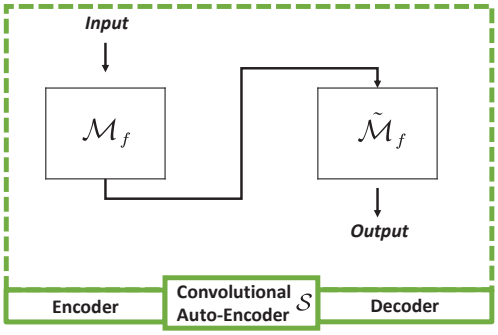

Fig. 6. The architecture of the convolutional auto-encoder $\mathcal{S}$.

TABLE I
KERNEL DESIGN OF CONVOLUTIONAL LAYERS

$\mathcal{M}_s$	$\mathcal{M}_f$	$\mathcal{G}$
Conv 7×7, 64		
ResBlock I $(\times 2)$ [1]	Conv 5×5, 32	**ResBottleneckBlock I** $(\times 2)$
Conv 3×3, 64	Conv 5×5, 64	Conv 1×1, 128
Conv 3×3, 64	Conv 3×3, 64	Conv 3×3, 128
ResBlock II $(\times 2)$	Conv 4×4, 128	Conv 1×1, 512
Conv 3×3, 128		
Conv 3×3, 128		

[1] The bold denotes the residual block or bottleneck block, while the number '2' behind the bold means the block repeats twice and the unbold below represents the detailed composition of each block.

then sent into the decoder $\tilde{\mathcal{M}}_f$ to reconstruct $\boldsymbol{\mathcal{P}_f^j}$. The corresponding output of $\tilde{\mathcal{M}}_f$ is denoted by $\tilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{f}}^{\boldsymbol{j}}$. We conduct the optimization of $\mathcal{S}$ by minimizing the mean square loss between the input of $\mathcal{M}_f$ and the output of $\tilde{\mathcal{M}}_f$. The loss function is then given by

$$Loss_{\mathcal{S}} = \frac{1}{|D_s|} \sum_{j=1}^{|D_s|} (\tilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{f}}^{\boldsymbol{j}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{f}}^{\boldsymbol{j}})^2. \tag{4}$$

Given the loss function $Loss_{\mathcal{S}}$, $\mathcal{S}$ is trained by using the stochastic gradient descent algorithm.

Let $\mathcal{M}_s$ and $\mathcal{M}_f$ reuse the parameters given by $\mathcal{R}$ and $\mathcal{S}$, respectively. Then, we train the SAR target classification network $\mathcal{G}$ on the source dataset, for which the cross-entropy loss function is applied and given by

$$Loss_{\mathcal{G}^s} = -\frac{1}{|D_s|} \sum_{j=1}^{|D_s|} \sum_{i=1}^{N_s} L_i(\boldsymbol{C_j}) \log_2(\tilde{L}_i(\boldsymbol{C_j})) \tag{5}$$

where $\tilde{L}_i(\boldsymbol{C_j})$ represents the i-th element of the output of $\mathcal{G}$. Given the loss function $Loss_{\mathcal{G}^s}$, $\mathcal{G}$ is trained by using the stochastic gradient descent algorithm.

B. Fine-tuning

In the fine-tuning stage, $\mathcal{M}_s$, $\mathcal{M}_f$, and $\mathcal{G}$ reuse the pre-trained parameters except for the last fully connected layer of $\mathcal{G}$, which is initialized by using a uniform distribution between $[0, 1]$. Then, the entire HNCSF is trained as a whole on the target dataset by fine-tuning $\mathcal{G}$. The cross-entropy loss function is applied and given by

$$Loss_{\mathcal{G}^t} = -\frac{1}{|D_t|} \sum_{j=1}^{|D_t|} \sum_{i=1}^{N_t} L_i(\boldsymbol{C_j}) \log_2(\tilde{L}_i(\boldsymbol{C_j})) \tag{6}$$

where $|D_t|$ represents the total number of SAR images of the target dataset.

Due to the inherent connection between the source and target datasets, the features extracted from the source dataset can be regarded as a more general representation of the few-shot SAR targets in the target dataset. Thus, the features extracted by the pre-trained networks are expected to be able to reflect the characteristics of the few-shot SAR targets.

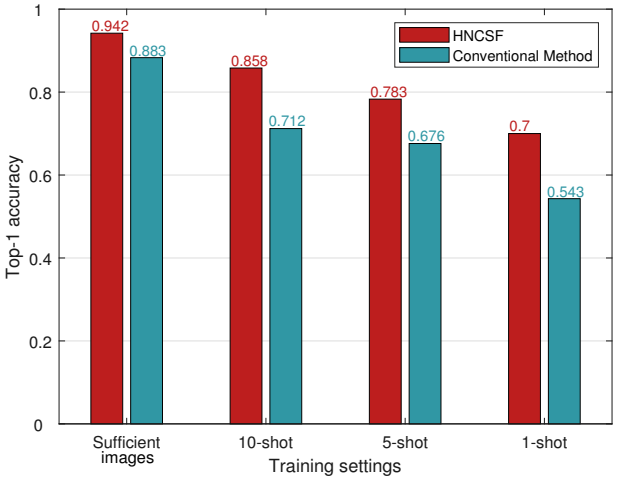

Fig. 7. Top-1 accuracies of the proposed HNCSF and the conventional method in different training settings.

Therefore, even if only a few images are available, a fine-tuning process can help the pre-trained HNCSF to identify the few-shot SAR targets.

V. Experimental Results

A. Experimental Settings

The SAR images in the S1 dataset are cropped to the size of 64×64. Table I shows the detailed kernel design of each convolutional layer of the proposed HNCSF. The trained HNCSF is tested on a dataset with 40 images per SAR target category. We compare the proposed HNCSF with a conventional method based on convolutional neural networks and model-agnostic meta-learning [16].

B. Performance

As shown in Fig. 7, our proposed HNCSF acquires the top-1 accuracy of 94.2% in the case that sufficient images are available. As we reduce the number of available images by constructing a 3-way 10-shot scenario, a 3-way 5-shot scenario, and a 3-way 1-shot scenario, the top-1 accuracy of our proposed HNCSF decreases but still maintains between 70.0-85.8%. The results indicate that our proposed HNCSF

can leverage the prior knowledge contained in the source dataset effectively to improve the target domain performance with only a few available samples.

Also, our proposed HNCSF outperforms the conventional method based on convolutional neural networks and model-agnostic meta-learning in the case of sufficient samples by 5.9%. In addition, the performance advantage of the proposed HNCSF increases in few-shot scenarios. Specifically, the top-1 accuracy improvements are 14.6%, 10.7%, and 15.7% for the 3-way 10-shot, 3-way 5-shot, and 3-way 1-shot scenarios, respectively. The superior performance comes from the additional frequency information, which provides extra features to assist the SAR target classification.

VI. Conclusion

In this paper, we proposed a hybrid classification network HNCSF for the few-shot SAR target classification. Specifically, we designed two feature extractors, which effectively extract the features from the spatial and frequency information of SAR images, respectively. Also, we adopted the transfer learning technology, which allows the HNCSF to extract prior knowledge contained in other related targets with a sufficiently large number of SAR images. Numerical results show that the proposed HNCSF maintains high top-1 accuracy in various few-shot scenarios, and outperforms the conventional method based on convolutional neural networks and model-agnostic meta-learning.

References

[1] S. Chen *et al.*, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806-4817, Aug. 2016.

[2] G. Dong and G. Kuang, "Classification on the monogenic scale space: Application to target recognition in SAR image," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2527-2539, Aug. 2015.

[3] W. Wu *et al.*, "High-resolution PolSAR scene classification with pre-trained deep convnets and manifold polarimetric parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6159-6168, Oct. 2018.

[13] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. ICML'17*, Sydney, Australia, Aug. 2017.

[4] S. Wagner, "Combination of convolutional feature extraction and support vector machines for radar ATR," in *Proc. IEEE FUSION'14*, Salamanca, Spain, Jul. 2014.

[5] Q. Guo and F. Xu, "A deep feature transformation method based on differential vector for few-shot learning," in *Proc. IEEE IGARSS'21*, Brussels, Belgium, Jul. 2021.

[6] W. Yu, T. Wang, and S. Wang, "Multi-label learning based antenna selection in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7255-7260, Jul. 2021.

[7] Z. Kuai and S. Wang, "Fair virtual network function scheduling with deep reinforcement learning," in *Proc. IEEE GLOBECOM'21*, Madrid, Spain, Dec. 2021.

[8] X. Guo, T. Wang, and S. Wang, "Joint optimization of caching and routing strategies in content delivery networks: A big data case," in *Proc. IEEE ICC'19*, Shanghai, China, May 2019.

[9] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 5, pp. 1019-1034, May 2015.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowledge Data Eng.*, vol. 22, no. 10, pp. 1345-1359, Oct. 2020.

[11] M. Huisman, J. Rijn, and A. Plaat, "A survey of deep meta-learning," *Artif. Intell. Rev.*, vol. 54, no. 6, pp.4483-4541, Apr. 2021.

[12] O. Vinyals *et al.*, "Matching networks for one shot learning," in *Proc. NeurIPS'16*, Barcelona, Spain, Dec. 2016.

[14] A. Santoro *et al.*, "Meta-learning with memory-augmented neural networks," in *Proc. ICML'16*, New York, NY, USA, Jun. 2016.

[15] W. Li *et al.*, "Few-shot SAR automatic target recognition based on Conv-BiLSTM prototypical network," *Neurocomputing*, vol. 443, no. 12, pp. 235-426, Jul. 2021.

[16] K. Fu *et al.*, "Few-shot SAR target classification via metalearning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1-14, Feb. 2021.

[17] R. Yang *et al.*, "Learning relation by graph neural network for SAR image few-shot learning," in *Proc. IEEE IGARSS'20*, Waikoloa, HI, USA, Feb. 2020.

[18] J. Souyris, C. Henry, and F. Adragna, "On the use of complex SAR image spectral analysis for target detection: Assessment of polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2725-2734, Dec. 2003.

[19] J. Singh and M. Datcu, "SAR target analysis based on multiple-sublook decomposition: A visual exploration approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 247-251, Mar. 2012.

[20] Z. Huang *et al.*, "Deep SAR-Net: Learning objects from signals," *ISPRS J. Photogramm. Remote Sens.*, vol. 161, pp. 179-193, Mar. 2020.

[21] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE CVPR'16*, Las Vegas, NV, USA, Jun. 2016.