A Domain Generalization Method for Heterogeneous SAR Image Target Classification

Siyuan Zhao
Comprehensive Training Department
Air Force Communication NCO Academy
drickzsy@163.com

Kaiming Li
Institute of Information and Navigation
Air Force Engineering University
Xi'an, China
kimysunrsp@163.com

Ying Luo
Institute of Information and Navigation
Air Force Engineering University
Xi'an, China
luoying2002521@163.com

Dan Xu
The Academy of Advanced Interdisciplinary Research
Xidian University
Xi'an, China
danxu@xidian.edu.cn

Jiajia Song
Comprehensive Training Department
Air Force Communication NCO Academy
Dalian, China
517102445@qq.com

Zhiwei Zhao
Comprehensive Training Department
Air Force Communication NCO Academy
Dalian, China
zhaozhiwei8566@126.com

Yong Kang
Comprehensive Training Department
Air Force Communication NCO Academy
Dalian, China
ky04608@qq.com

***Abstract*—Different imaging platforms and imaging conditions will lead to inconsistent distributions of Synthetic Aperture Radar (SAR) images, resulting in heterogeneous SAR images. In order to solve the problems of inconsistent distribution of heterogeneous SAR image data and lack of generalization in model training, this paper proposes a novel domain generalization target classification method. First, we optimize the features of each source domain to prepare prompts for finding the best direction. Then, we train two embedded networks to mix the expert prompt appropriately so that the output achieves the best direction. Extensive comparative experiments demonstrate that the proposed domain generalization method exceeds the existing state-of-the-art methods in terms of classification accuracy.**

***Index Terms*—Synthetic Aperture Radar, target classification, source domain, domain generalization**

I. INTRODUCTION

Synthetic Aperture Radar (SAR) is a type of active microwave imaging sensor that can observe the Earth in all-weather and all-day conditions [1]. SAR image target classification is an important part of the automatic target recognition task [2]. The use of SAR image data for target classification in the field of deep learning has become an important research direction. Compared with optical data, the amount of SAR image data is relatively scarce at this stage, and it is difficult to train a model with good recognition performance and high robustness using only the existing SAR image data [3; 4].

Most of the existing deep learning-based target recognition methods have achieved optimal performance when both the training and test sets come from SAR data with the same joint probability distribution [5]. However, most of the existing deep learning-based studies do not directly address the pain points of the SAR image target classification task, i.e., the inconsistent distribution of the source data and the generalization of the algorithms themselves. With the successive release of SAR imaging platforms around the world, more and more SAR image data have been used to carry out algorithm research and validation. Existing research on SAR target recognition rarely considers the sensitivity of imaging results caused by variations in specific SAR imaging conditions, including changes in sensor parameters, imaging modes, and differences in target scenes. [6].

In real-world scenarios, models are usually expected to fit multiple target domains (test data) and be able to perform fast small-sample fine-tuning, and training and test are usually performed separately, so that the trained model may have to be tested in another unfamiliar scenario every once in a while, which results in the data used for training in order to debug the model being unavailable during the test phase [7]. With proper network parameters, traditional deep learning models iterate over the training data several times to achieve a full fit, while most existing models assume that the training data and testing data are independently and identically distributed, and when faced with an unknown distribution of the data to be tested, the robustness of the existing model will be challenged and its credibility will be reduced. In this paper, we propose a novel domain generalization method that introduces expert cues to optimize the source domain features in order to guide the training of two embedded network models so as to efficiently merge these expert cues to obtain the best output.

In the following, the details of the algorithm are unfolded in detail.

II. METHODOLOGY

Domain generalization is a task that typically uses a known source domain to fit a model to an unseen target domain. First, it performs domain adaptation by designing "expert" specific adaptation cues for each source domain. This step is processed end-to-end mainly through backpropagation. A subsequent phase dedicated to domain generalization creates target domain-specific sample cues for each input image by determining the average of the weights of all experts through an attention-based algorithm. In this phase, the system uses two separate trainable encoders: one for the input images and the other for the pre-trained experts. The expert's weights are derived from encoding the similarity between the input image sample and the expert embedding. These two phases are called "Source Domain Feature Optimization" and "Expert Prompt to Optimize Joint Training", respectively.

In this section, the proposed method is presented in detail. In our paper, we define source domain dataset as $\mathcal{S}_{j\in[1,M]} = \{x_i^s, y_i^s\}_{i=1}^{N_s}$, where x_i^s is ith image in N_s fully labeled SAR images, y_i^s denotes the corresponding category label of x_i^s. Meanwhile, there are K categories of ships in the dataset, i.e. $y_i^s \in \{1, 2, \ldots, K\}$. There are M differently distributed data domains to act as source domains during the training of the algorithm, while unseen target domain is introduced at the time of testing to validate the performance of the algorithm. The overall framework of proposed method is shown in Fig. 1.

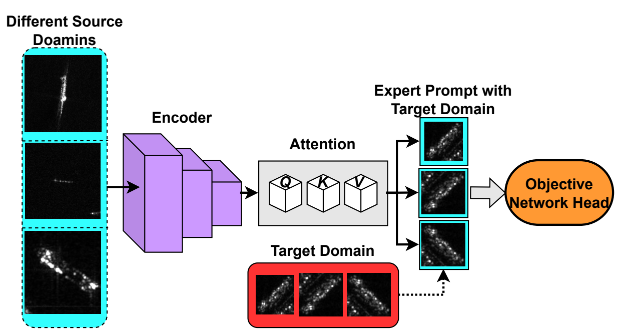

Fig. 1. The proposed domain generalized SAR image target classification network structure. The source domain is a number of differently distributed SAR image datasets. The source domain inputs are optimized with network features to obtain expert prompt, which is then combined with the target domain data for joint training for target classification.

A. Source Domain Feature Optimization

Assume that the expert in the jth domain is $p_j \in R^{d_{prompt}}$, where d_{prompt} is the dimension of the prompt. Then, $p_j \in [1, N]$ represents the optimal direction to move the input in the source domain, which we can optimize using the known source sample. Prompt for the target domain $p_{\mathcal{T}}$ cannot be optimized directly because target domain is invisible.

We approximate $p_{\mathcal{T}}$ as a linear combination of $p_j \in [1, N]$ as follows

$$p_{\mathcal{T}} = \sum_{j=1}^{M} \lambda_j p_j, \tag{1}$$

where $\lambda_j = \Lambda\left(p_j | \mathbf{x} \in \mathcal{S}\right)$ and Λ is a conditional function representing the optimal weights of p_j given $\mathbf{x} \in \mathcal{S}$.

Next, the following variable Kullback-Leibler-divergence is introduced as the objective function

$$\mathcal{L}_1 = \mathrm{JS}\left(\aleph\left(x_{\mathcal{T}} + p_{\mathcal{T}}\right) || \mathcal{D}_{\mathcal{T}}\right), \tag{2}$$

where $x_{\mathcal{T}} \in \mathcal{T}$, $\aleph$ is the whole network framework and $\mathcal{D}_{\mathcal{T}}$ is the target sample distribution of $x_{\mathcal{T}} + p_{\mathcal{T}}$. JS is Jensen–Shannon-divergence which is a special form of the Kullback-Leibler-divergence.

By iteratively optimizing this objective function, the following inference can be drawn. If there exists a range of optimal tips for each domain, then the expert must be a point within the range. The objective is to blend multiple expert tips into one tip. To do this, each expert must be proficiently trained in their primary domain (in our case, the domain). We utilize adversarial reprogramming, a direct gradient-based approach, to tune these experts. While this approach is sufficient in specific cases, it is not sufficient in domains that differ significantly from the pre-trained domain. To address this problem, we use meta-prompt to initialize the expert prompt, which is to build self-improving agent's tips with a generic prompt. Meta-prompt refer to pre-trained prompt that can be used to initialize visual prompt.

B. Expert Prompt to Optimize Joint Training

A further idea is to combine experts so that images from unseen domains are correctly classified. We combine the experts by weighted average. The weights must indicate how many experts are required for a particular image. This requirement can be realized by a cross-focusing mechanism. In this case, the experts become the keys (K) and values (V), and the target image becomes the query (Q) of interest. We use embedding vectors instead of directly comparing Q and K. We have a pre-trained network as a shared embedding network with two different trainable head linear layers for each of Q and K. Q and K are embedded via a shared encoder and their respective linear heads. A scalar attention weight equal to the number of experts is then obtained, and VQK^{T} becomes the cue for the target image.

The pairwise manipulation of features using only such an attention mechanism will inevitably run into problems of scale differences and weight saturation. First, the experts are optimizing independently in different domains, which results in significant differences in scale. Second, since the weights are computed independently, without the use of scaling tools such as softmax functions, the weights may become over-saturated.

The countermeasure we propose here is that each expert is normalized with its own L2 regularization value. Simultaneously, the tanh function is used to map the weights onto the new interval $[-1,1]$, thus avoiding saturation to some extent. In this way, the expert prompt for the ith sample of the target domain can be redefined in the following form

$$p_{\mathcal{T},i} = \sum_{j}^{M} \frac{p_j}{\|p_j\|^2} e_{\mathcal{T}}\left(x_{\mathcal{T}}, i\right) e_{\mathcal{E}} \left(\frac{p_j}{\|p_j\|^2}\right)^{\mathrm{T}}, \tag{3}$$

where $e_{\mathcal{T}}$ and $e_{\mathcal{E}}$ are the embeddings for target samples and experts, respectively. When the generalization training is completed, the expert embedding vectors are fixed because the experts do not change. Therefore, the expert embedding procedure is no longer needed in the evaluation.

III. EXPERIMENTS

A. Datasets and Experiment Settings

In this section, we perform out-of-domain evaluation experiments to demonstrate the effectiveness and superiority of the proposed method, mainly on SAR image datasets with different distributions. Three typical satellite-borne SAR image data are selected for the experimental procedure, they are TerraSAR-X, Sentinel-1 and Gaofen-3. As the parameters of their imaging platforms, imaging conditions and external environment are somewhat different, which makes the image data from these three datasets have different statistical distributions, which can be modeled by designing the proposed algorithm to improve the generalization performance. Of course, due to the large number of target types collected by the group, here, in order to unify the target types, we screen the four types of ship targets in these three datasets as the experimental objects, which are Cargo, Fishing, Tanker and Others.

There are obvious differences among these three satellites, and Fig. 2 shows the corresponding four categories of ships in these datasets. It is clear that the corresponding target images presented in the three datasets have large differences, which can lead to overfitting of the model if the supervised learning strategy is extended.

Empirically, ResNet-101 is employed as the backbone network whose weights are pre-trained on ImageNet. Stochastic Gradient Descent is chosen with a momentum of 0.9 as the optimization method for the training process. The initial learning rate is set to 1×10^{-4}. All of the experiments are performed on a GIGABYTE X299WU8 workstation with an Intel Core(R) i9-10920X CPU, an Nvidia TESLA V100 Turbo GPU, and 32G RAM under Centos 7. Our code is constructed on the ResNet-101 backbone network based on the Pytorch framework.

B. Results and Analysis

In order to demonstrate the superiority and accuracy of the proposed method, the proposed method is compared with the existing state-of-the-art (SOTA) methods in the comparative experiment. Typical of several existing SOTA methods include: Towards Recognizing Unseen Categories in Unseen Domains (TRUC) [8], Exploiting Domain-Specific Features to Enhance Domain Generalization (EDSF) [9], Domain-invariant Feature Exploration for Domain Generalization (DIFE) [10] and Domain-Specific Risk Minimization for Domain Generalization (DSRM) [11]. With three data domains consisting of three different satellite-borne SAR datasets, three different domain generalization tasks can be designed in the experiment, i.e., two of them are used as source domains for model generalization training, while the other is used as the target domain to be tested.

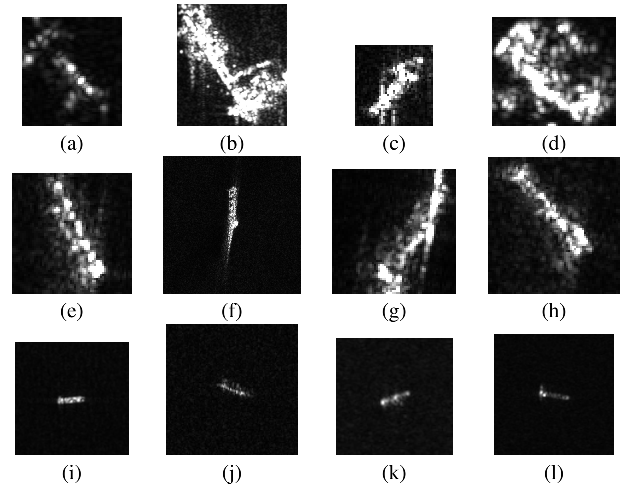

Fig. 2. Four categories of ships in the three datasets. The first row is from TerraSAR-X, the second row is from Gaofen-3 and the third row is from Sentinel-1. (a), (e) and (i) Cargo; (b), (f) and (j) Fishing; (c), (g) and (k) Tankers; (d), (h) and (l) Others.

Thus, the experimental results are shown in Table I, Table II and Table III, respectively. The test process entails calculating the classification accuracy and mean accuracy (mAP) for each category. From these three tables, the proposed method is superior to all SOTA methods among three tasks.

Although the experimental results for domain generalization in target classification show relatively low accuracy values in the tables, they gradually demonstrate these methods' model generalization capability. Notably, our proposed method consistently outperforms all SOTA methods across all three tasks. The experts are normalized so that they initially have the same scale through the normalization in Eq. (3). Since each weight is computed independently of other experts, attention weights can be saturated during training. Thus, all these processes lead to the optimization of the performance of the proposed method.

TABLE I
TARGET CLASSIFICATION RESULTS OF GENERALIZED MODEL TRAINED WITH TERRASAR-X AND GAOFEN-3 AS SOURCE DOMAINS ON SENTINEL-1 AS THE TARGET DOMAIN

	Cargo	Fishing	Tanker	Others	mAP
TRUC	64.6	22.5	36.8	44.5	42.1
EDSF	65.2	25.6	37.4	46.2	43.6
DIFE	65.7	26.9	38.6	49.6	45.2
DSRM	67.4	27.1	38.9	50.0	45.9
Ours	**68.9**	**29.8**	**39.7**	**55.6**	**48.5**

TABLE II
TARGET CLASSIFICATION RESULTS OF GENERALIZED MODEL TRAINED WITH TERRASAR-X AND SENTINEL-1 AS SOURCE DOMAINS ON GAOFEN-3 AS THE TARGET DOMAIN

	Cargo	Fishing	Tanker	Others	mAP
TRUC	64.6	28.8	38.8	59.0	47.8
EDSF	65.2	29.9	41.4	61.5	49.5
DIFE	65.7	30.7	42.6	**61.8**	50.2
DSRM	67.4	32.1	44.7	59.0	50.8
Ours	**68.9**	**33.4**	**44.9**	59.6	**51.7**

TABLE III
TARGET CLASSIFICATION RESULTS OF GENERALIZED MODEL TRAINED WITH SENTINEL-1 AND GAOFEN-3 AS SOURCE DOMAINS ON TERRASAR-X AS THE TARGET DOMAIN

	Cargo	Fishing	Tanker	Others	mAP
TRUC	71.6	31.8	43.8	68.0	53.8
EDSF	72.2	33.9	44.1	67.8	54.5
DIFE	73.3	34.2	45.6	67.7	55.2
DSRM	74.8	35.1	46.0	71.3	56.8
Ours	**75.9**	**36.1**	**47.9**	**74.9**	**58.7**

IV. CONCLUSIONS

With an increasing number of SAR imaging platforms emerging in this field, there is growing urgency for categorizing heterogeneous SAR image data effectively. In order to solve the problems of inconsistent data distribution and poor model generalization performance in the classification process of heterogenous SAR image data, this paper proposes a new domain generalization target classification method. First, we optimize the features of each source domain to prepare the prompt for finding the best directions. Then, we train two embedded networks to appropriately mix the expert prompt so that the output achieves the best direction. Comparative experiments demonstrate the high classification accuracy of the proposed method over the SOTA methods. In the future, we should be more committed to researching the integration of domain generalized SAR image classification methods into remote sensing LLM, and how to improve the quality and accuracy of the output of large language models is one of the key points for applications in various fields.

ACKNOWLEDGMENT

This work is supported in part by National Natural Science Foundation of China under grant 62401620, 62131020 and 62401430.

REFERENCES

[1] Andreas Theodosiou and Paco López-Dekker, "On the Sensitivity to Height and Motion of Bistatic SAR Interferometry: A Spectral View," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–12, 2024.

[2] Shengkai Sun, Zi He, Zhenhong Fan, and Dazhi Ding, "SAR Image Target Recognition Using Diffusion Model and Scattering Information," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[3] Yu Shi, Lan Du, Yuchen Guo, Yuang Du, and Yiming Li, "Unsupervised domain adaptation for ship classification via progressive feature alignment: From optical to sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–17, 2024.

[4] Siyuan Zhao, Ying Luo, Tao Zhang, Weiwei Guo, and Zenghui Zhang, "A Domain Specific Knowledge Extraction Transformer Method for Multisource Satellite-borne SAR Images Ship Detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 198, pp. 16–29, 2023.

[5] Siyuan Zhao, Yuanpeng Zhang, Ying Luo, Yong Kang, and Hui Wang, "Dynamically Self-Training Open Set Domain Adaptation Classification Method for Heterogeneous SAR Image," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[6] Siyuan Zhao, Zenghui Zhang, Weiwei Guo, and Ying Luo, "An Automatic Ship Detection Method Adapting to Different Satellites SAR Images With Feature Alignment and Compensation Loss," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[7] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip S. Yu, "Generalizing to Unseen Domains: A Survey on Domain Generalization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052–8072, 2023.

[8] Massimiliano Mancini, Zeynep Akata, Elisa Ricci, and Barbara Caputo, "Towards Recognizing Unseen Categories in Unseen Domains," *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII*, p. 466–483, 2020.

[9] Manh-Ha Bui, Toan Tran, Anh Tran, and Dinh Phung, "Exploiting Domain-Specific Features to Enhance Domain Generalization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21189–21201, 2021.

[10] Wang Lu, Jindong Wang, Haoliang Li, Yiqiang Chen, and Xing Xie, "Domain-invariant Feature Exploration for Domain Generalization," *Transactions on Machine Learning Research*, 2022.

[11] Yifan Zhang, Jindong Wang, Jian Liang, Zhang Zhang, Baosheng Yu, Liang Wang, Dacheng Tao, and Xing Xie, "Domain-Specific Risk Minimization for Domain Generalization," *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3409–3421, 2023.