DISCOVERING NOVEL CATEGORIES IN SAR IMAGES IN OPEN SET CONDITIONS

Liu Dai[1,4], *Weiwei Guo*[2*], *Zenghui Zhang*[3], *Wenxian Yu*[3]

[1]College of Electronic and Information Engineering, Tongji University, Shanghai, 200092, China
[2]Center of Digital Innovation, Tongji University, Shanghai, 200092, China
[3]Shanghai Key Laboratory of Intelligent Sensing and Recognition, Shanghai Jiao Tong University, Shanghai, 200240, China
[4]LunarAI, Shanghai, 200439, China

ABSTRACT

In this paper, we deal with the issue of discovering data of novel categories for Synthetic Aperture Radar (SAR) images under open set conditions. The traditional SAR image classification methods are trained under the close-set setting where all categories in testing data are seen in training data. It does not always meet the requirements of the real SAR imagery interpretation applications. With a labelled SAR image dataset, we propose a multi-stage approach to effectively pick out images belonging to new classes in another unlabelled dataset and then cluster them into correct number of novel categories. To do so, our pipeline is composed of three major steps: (1) train a powerful feature extractor leveraging both the labelled and unlabelled dataset by semi-supervised inference; (2) identify the unknown data by openset detection; (3) cluster these unknown data based on the features generated by the extractor to discover novel categories. The proposed method is validated on a Sentinel-1 SAR image dataset OpenSARUrban [1].

Index Terms— SAR, Image Classification, Open Set Recognition, Novel Category Discovery, Clustering

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is of prominent importance for Earth observation due to its capability of operation in all weather and day-and-night conditions. There is an increasing amount of SAR image data collected and publicly accessible. However, it is still very challenging for SAR image interpretation because of its special active microwave imaging mechanism. Currently, deep neural networks (DNNs), e.g., convolutional neural networks (CNNs), have been widely applied for various SAR image interpretation tasks[2, 3, 1]. Previous methods only work well for close-set scenarios. With a large amount of SAR data, it is very difficult to annotate the data of all categories, which requires expertise knowledge and consumes much time and labour resource. However, a new unlabelled dataset may contain new data of novel classes that does not appear in training data. Accordingly, it is of practical use to come up with an approach to identify images within a new SAR dataset not belonging to the categories we have known and discover their categories.

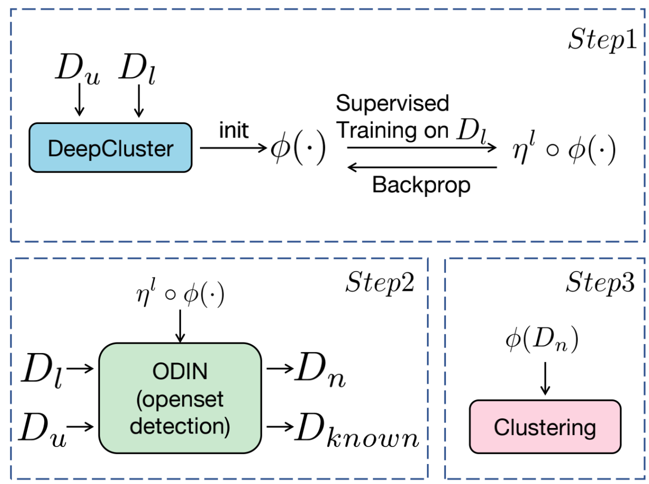

Fig. 1. The detailed pipeline of our method to pick out the new category data in SAR image dataset, conducting Step 1,2 and 3 in turn.

Some related work on **Novel Category Discovery (NCD)**, **Openset Detection** and **Representation Learning** can inspire and help us to achieve our goal. Recently, Han *et al.* [4] formalized the issue of NCD and proposed a pipeline based on prior knowledge learned by a different but relative labelled dataset. Han *et al.* [5] also explored using ranking statistics to measure the similarity between images and proposed a method estimating the number of classes in an unlabelled dataset. Zhao *et al.* [6] further ameliorated the work with dual ranking statistics and mutual knowledge distillation. Liang *et al.* [7] proposed an effective baseline for out-of-distribution data detection without requirements for pre-trained neural networks. Caron *et al.* [8] addressed the

This work was supported in part by the National Natural Science Foundation of China under grant 62071333,and ESA-MOST CHINA Dragon 5 programm ID.58190.

issue of learning useful low-level visual features in a self-supervised way by assigning clustering results as pseudo labels.

Considering that characteristics of SAR images differ considerably from common optical RGB images which often contain speckle and slant-range distortions, therefore in our pipeline, a good feature extracted from raw data is the key foundation to distinguish images from different categories and to all the operations coming after. We design a three-stage method to firstly learn a suitable feature extractor via semi-supervised learning and then conduct openset detection, estimating the number of novel classes and then clustering all based on the feature generated by the extractor. Compared with the previous methods, our proposed method can not only detect the unknowns but also discover new categories.

2. METHODOLOGY

The pipeline of our proposed method is shown in fig. 1. Given a labelled known dataset $D_l = \{(x_i^l, y_i^l); i = 1, 2, ..., N\}$ with the number of classes C_l and an unknown dataset $D_u = \{x_i^u; i = 1, 2, ..., M\}$ with the number of classes C_u, where $x_i \in \mathbb{R}^{1\times H\times W}$, our aim is to identify a subset D_n within D_u which belong to novel categories with the number of classes $C_n = C_u - C_l$. The pipeline consists of three major steps detailed below.

2.1. Feature Extractor

Now we have two datasets D_l and D_u, whose categories are known and unknown respectively. Some works have shown that tasks like clustering or openset detection can perform better in the appropriate high-dimensional feature space [4, 5, 6]. Accordingly, we put forward a two-stage method with CNN architecture to roll out 'a good' feature extractor.

Self-supervised Learning To avoid the feature extractor being only sensitive to the labelled data, we use both D_l and D_u to bootstrap the feature extractor $\phi : x \to \phi(x) \in \mathbb{R}^d$ via self-supervised training. Several self-supervised methods have been proposed such as RotNet [9] and MoCo[10]. Given that SAR images are sensitive to imaging operation conditions and the massive number of SAR images require a great amount of resource if we take some contrastive learning methods, we thus use DeepCluster [8] to initialize ϕ with the VGG16 [11] architecture. In detail, we iteratively conduct K-Means [12] on $\phi(D_l \cup D_u)$ and assign the clustering results as pseudo labels. In each iteration, we extend ϕ with a classification head η^p: $\mathbb{R}^d \to \mathbb{R}^k$, where a linear layer is implemented to classify the input data into K classes. We use the standard cross-entropy loss:

$$L_{DeepCluster} = -\frac{1}{N+M}\sum_{1}^{N+M}\log(\eta^p \circ \phi(x_i))_{l_p}$$

Table 1. The scores for clustering data of 5 classes with the feature extractor before and after being fine-tuned on D_l, where ACC refers to Clustering Accuracy, NMI refers to Normalized Mutual Information and ARI refers to Adjusted Rand Index.

Feature Extractor	Number of Clustering Class	ACC	NMI	ARI
before fine-tuning	5	29.54%	0.127	0.110
after fine-tuning	5	43.22%	0.220	0.167

where $x_i \in D_l \cup D_u$, l_p represents the pseudo labels assigned after each clustering.

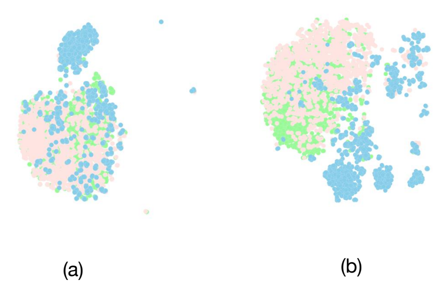

Fig. 2. Visualization of the distribution of features extracted from D_n by ϕ. (a) features before fine-tuning; (b) features after fine-tuning.

Supervised Fine-tuning Now we have the feature extractor ϕ trained by DeepCluster [8] on the union of D_u and D_l. However, self-supervised training ignores all labels and therefore cannot leverage the information contained in $\{x_i^l, y_i^l\} \in D_l$. To address the problem, we further extend ϕ with a new classification head η^l: $\mathbb{R}^d \to \mathbb{R}^l$, where we implement a linear layer and a softmax layer coming after. The function $\eta^l \circ \phi(\cdot)$ is fine-tuned on the labelled dataset D_l in a supervised way to learn a classifier for C_l known classes. Here we take a standard cross-entropy loss defined as:

$$L_{supervised} = -\frac{1}{N}\sum_{1}^{N}\log(\eta^l \circ \phi(x_i))_{y_i}$$

where $(x_i, y_i) \in D_l$.

The distribution of features extracted by ϕ before and after fine-tuned are visualized in fig. 2(a) and fig. 2(b), where we can see $\phi(x)$ distributes more separately after being supervised-trained on D_l. We also validate the effectiveness of fine-tuning by experiments shown in table 1, showing that clustering based on $\phi(D_n)$ achieves a better performance

Table 2. The scores for clustering the novel dataset D_n.

Data Process	Data Distribution (known+unknown)	ACC	NMI	ARI
$\phi(D_n)$	7 + 3	58.94%	0.310	0.272
$\phi(D_n)$	6 + 4	53.55%	0.256	0.169
$\phi(D_n)$	5 + 5	43.22%	0.220	0.167
raw D_n	7 + 3	41.52%	0.127	0.058

after fine-tuning ϕ on D_l and self-supervised learning alone cannot effectively leverage information for downstream tasks.

2.2. Openset Detection for Novel Data

With the trained feature extractor, we now are ready to pick out the unknown data within D_u that are novel to D_l in the new feature space. With the function $\eta^l \circ \phi(\cdot)$, we use ODIN [7] based on our CNN network to conduct the openset detection, setting D_l as in-distribution images and D_n as out-of-distribution images.

In detail, we adjust the softmax layer in η^l with temperature scaling T [7] and for each input we preprocess the image by adding small perturbations [7] before sending the image into our own CNN classifier $\eta^l \circ \phi(\cdot)$. In the end, we recognize each input image from D_u as in- or out-of-distribution data by calculating the new softmax score.

2.3. Clustering on Novel Data

Now we have the novel dataset D_n consisting of the unknown data in D_u. But before generating them into classes, we need to figure out the number of categories C_n for D_n. With the feature extractor ϕ we can estimate C_n and set it as K to finally cluster D_n into correct categories by K-Means.

Estimate the number of K: A light but effective method to estimate K on some novel dataset was proposed in [4]. We set our D_l as the probe set mixing with D_n and conduct K-Means on the features extracted by ϕ. By evaluating ACC for the subset D_l^{rv} of D_l and CVI for D_n as defined in [4], we precisely estimate K of D_n, using our own feature extractor ϕ in the method proposed in [4].

Clustering: K-Means is conducted on $\phi(D_n)$ instead of directly on the raw data D_n, achieving favorably both in clustering effect and in efficiency.

3. EXPERIMENTS

3.1. Dataset

We utilize a Sentinel-1 SAR image dataset OpenSARUrban [1] to test our pipeline, which provides 33,358 patches of images covering ten different classes, VH or VV polarized. We only retain images VH-polarized and divide them into D_l and D_u in the ratio of 5:5. And, D_l covers 7 classes whereas D_u covers the total 10 classes, including the 7 classes contained in D_l. For comparison, we also test the pipeline with the number of classes in D_l being 6 and 5.

3.2. Training details

For DeepCluster stage in Section2.1, we set K as 512 and the network parameters initialized to the original VGG16 network. The training is on the union of D_l and D_u for 100 epochs. fig. 3(a) shows the evolution of Normalized Mutual Information(NMI) on the previous assignment of pseudo labels after each loop, defined as:

$$NMI(Y, \bar{Y}) = \frac{I(Y, \bar{Y})}{\sqrt{H(Y)H(\bar{Y})}}$$

where I measures the mutual information between assigned labels Y and predicted labels $\bar{Y}$, with H the entropy. We can see that NMI increases during our training and it means the features we extract progressively capture information related to object classes during the training.

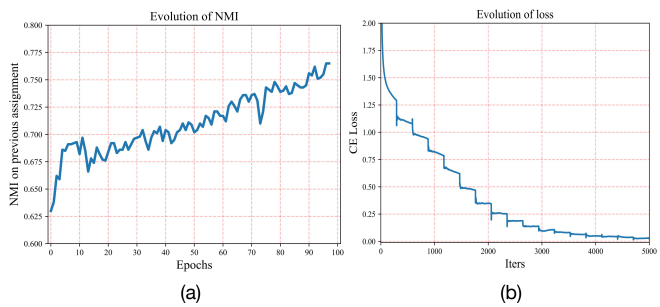

Fig. 3. The training process of self-supervised learning and supervised learning.

When fine-tuning the feature extractor ϕ along with the classification head η^l, we optimize the network with standard cross-entropy loss for 5000 iterations, updating the weights both in ϕ and η^l. The process is shown in fig. 3(b). For step2 and step3 in Fig.2, we keep the setting identical to the original method except for using our own CNN feature extractor ϕ and classification head η^l, keeping true positive rate (TRP) as 95% in ODIN for step2.

3.3. Results and Discussion

We estimate the number of categories for the novel dataset D_n and then conduct K-Means on $\phi(D_n)$. For comparison, we also test K-Means directly on D_n and find our methods outperforms it after previously extracting the feature with ϕ. The results are reported in table 2, where we can see clustering performs better on the features extracted by our ϕ, considering the extractor ϕ is also fundamental to openset detection and estimating the number of novel categories. We measure the clustering performance on D_n by the indexes of Accuracy (ACC), Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) [13], the last of which is defined as:

$$ARI = \frac{RI - E(RI)}{\max(RI) - E(RI)}$$

with

$$RI = \frac{TP + TN}{TP + FP + TN + FN}$$

where TP,TN,FP,FN refer to True Positive, True Negative, False Positive, False Negative respectively.

Also, we find the method proposed in [4] to estimate K can work with a lower error rate when using our feature extractor, where we can estimate the number of novel categories D_n accurately to be 3,4 and 5, data distribution as described in table 2.

In this work, we mainly focus on the pipeline itself that can effectively extract the features of SAR images and help discover the new categories. Accordingly, we choose the classical but simple architecture VGG16 as our backbone, instead of trying some more advanced and complex networks.

4. CONCLUSION

In this paper, we roll out a multi-stage pipeline to find out data with novel categories in unlabelled SAR dataset and cluster them into correct number of categories, based on an effective feature extractor learned from both labelled and unlabelled data. We conduct validation experiments on a Sentinel-1 SAR image dataset and have proven our method to be effective.

5. REFERENCES

[1] Juanping Zhao, Zenghui Zhang, Wei Yao, Mihai Datcu, Huilin Xiong, and Wenxian Yu, "Opensarurban: A sentinel-1 sar image dataset for urban interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020.

[2] Juanping Zhao, Weiwei Guo, Zenghui Zhang, and Y. U. Wenxian, "A coupled convolutional neural network for small and densely clustered ship detection in sar images," *Science China Information Sciences*, vol. 62, no. 004, pp. 42301, 2019.

[3] Lanqing Huang, Bin Liu, Boying Li, Weiwei Guo, Wenhao Yu, Zenghui Zhang, and Wenxian Yu, "Opensarship: A dataset dedicated to sentinel-1 ship interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 1, pp. 195–208, 2018.

[4] Kai Han, Andrea Vedaldi, and Andrew Zisserman, "Learning to discover novel visual categories via deep transfer clustering," in *ICCV*, 2019.

[5] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman, "Autonovel: Automatically discovering and learning novel visual categories," *IEEE TPAMI*, 2021.

[6] Bingchen Zhao and Kai Han, "Novel visual category discovery with dual ranking statistics and mutual knowledge distillation," in *NeurIPS*, 2021.

[7] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," *arXiv preprint arXiv:1706.02690*, 2017.

[8] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 132–149.

[9] Spyros Gidaris, Praveer Singh, and Nikos Komodakis, "Unsupervised representation learning by predicting image rotations," in *ICLR*, 2018.

[10] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020.

[11] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[12] James MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.

[13] Lawrence Hubert and Phipps Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.