IMPROVING GENERALIZATION FOR FEW-SHOT REMOTE SENSING CLASSIFICATION WITH META-LEARNING

Surbhi Sharma[1,2], *Ribana Roscher* [3,4], *Morris Riedel*[2], *Shahbaz Memon*[1], *and Gabriele Cavallaro*[1]

[1] Jülich Supercomputing Centre, Forschungszentrum Jülich, Germany
[2] School of Engineering and Natural Sciences, University of Iceland, Iceland
[3] Institute of Geodesy and Geoinformation, University of Bonn, Germany
[4] Data Science in Earth Observation, Technical University of Munich, Germany

ABSTRACT

In Remote Sensing (RS) classification, generalization ability is one of the measure that characterizes the success of Machine Learning (ML) models, but is often impeded by the scarse availability of annotated training data. Annotated RS samples are expensive to obtain and can present large disparities when produced by different annotators. In this paper, we utilize Few-Shot Learning (FSL) with meta-learning to address the challenge of generalization using limited amount of training information. The data used in this paper is leveraged from different datasets that have diverse distributions, that means distinct feature spaces. We tested our approach on publicly available RS benchmark datasets to perform few-shot RS image classification using meta-learning. The results of the experiments suggest that our approach is able to generalize well on the unseen data even with limited number of training samples and reasonable training time.

Index Terms— Few-shot learning, meta-learning, deep learning, classification, remote sensing.

1. INTRODUCTION

To achieve sufficient generalization performance in different RS classification problems with Deep Learning (DL) models, it is essential to have access to large amounts of annotated training samples with diverse data distribution [1]. However, such large amounts of annotated training samples are difficult to obtain. Training only on a single dataset (collection of related data) where training and validation data follow the same distribution decreases the generalization capability of the classifier. As a result, trained classifiers achieve better results on the test data of the same distribution than on samples of a different distribution, and thus are generally restricted to similar areas in the world.

Besides that, to rapidly discover and represent new classes with few observations and with high generalization capability, Few-Shot Learning (FSL) can be adapted with meta-learning paradigm [2]. FSL is a ML paradigm for supervised learning from a limited number of samples [3]. Together with meta-learning, they tackle meta-problems where the learning algorithms learn how to learn new classes from a limited number of training samples. The learning process becomes faster as they turn to the previous tasks for accumulated meta-knowledge. There are several categories for few-shot classification based on meta-learning such as distance-based approaches [4], optimization-based method [5], class or task-wise network modulation with amortization [6].

The main objective of this paper is to increase the generalization capability of ML and DL models for few-shot RS classification using meta-learning. The main contributions are: (1) a meta-dataset workflow [1] adapted from an existing pipeline [2] to create a more realistic classification approach for RS images (i.e., this pipeline can input different datasets that are acquired by distinct sensors) and (2), a distance-based meta-learning approach applied for few-shot RS classification [4] using the meta-dataset. The experimental results demonstrate that our approach can generalize on unseen RS data with a reasonable training time.

2. RECENT WORKS

Most works in the direction of FSL in RS include meta-learning approaches such as the utilization of a meta-learning model for scene recognition in RS images [7] for lifelong learning. To serve lifelong FSL, the knowledge is transfered from one dataset to another, with just a few samples. For few-shot scene classification in RS, approaches using meta-agnostic ML has been applied with promising results [8]. The authors of [9] introduce an inductive transfer learning

Part of this work was performed in the ADMIRE project, which received funding from the European Union's Horizon 2020 JTI-EuroHPC research and innovation programme under the grant agreement no. 956748. The authors gratefully acknowledge the computing resources from the DEEP-EST project, which received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no. 754304. The work is partly funded by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (Grant number: 01DD20001).

[1] `https://gitlab.jsc.fz-juelich.de/sdlrs/few-shot-remote-sensing-classification-with-meta-learning`

[2] `https://github.com/google-research/meta-dataset/tree/arxiv_v1`

problem where few data samples from a single region allow a model to adapt to an unseen region. They used the technique to evaluate the model agnostic meta-learning algorithm on classification and segmentation tasks using globally and regionally distributed datasets. In [10], a few-shot scene classification scheme is proposed. The authors combine multiple attention mechanisms and the attention-reference mechanism into the deepEMD network for few-shot classification on benchmark RS datasets. More recent studies considering few-shot classification mainly address classification based on meta-learning by prediction and finetuning the weights of the classifiers for new classes or methods that learn to augment data such as generative adversarial networks (GANs) [11].

3. METHODOLOGY

We consider two methods for few-shot RS classification: The first method is a meta-learning based approach and the other is a baseline approach without meta-learning (for comparison purposes).

3.1. Meta-learning approach to FSL

Meta-learning methods include two stages: **meta-learning and meta-testing stage**: In the meta-learning stage, as shown in Fig.1 (a), the meta-learner trains the learner on a training set that contains a large number of different classes. In so-called episodes, the meta-learning model is trained end-to-end with support sets (subsets of the training data) in order to make the classifier learn from the few samples. In this phase, the model learns common feature representations of all classes based on the acquired meta knowledge. In case of a new class, the model is fine-tuned with a small class-specific training data set (support set from the test episode) and evaluated using a query set.

In FSL, the model is able to generalize for a given set of episodes having C classes and few labeled samples per class ($k_l, l \in \{1, \ldots, C\}$), when trained on a support set $\mathcal{S}_s = \{(\boldsymbol{m}_1, n_1)...(\boldsymbol{m}_K, n_K)\}$, where $K = \sum_l k_l$, and tested on a query set $\mathcal{Q}_s = \{(\boldsymbol{m}_1^*, n_1^*)...(\boldsymbol{m}_K^*, n_K^*)\}$, where $\boldsymbol{m}$ is the input vector ($\boldsymbol{m} \in \mathbb{R}$) and n is the label ($n \in \{1, \ldots, C\}$). Here, $\boldsymbol{m}^*$ and n^* are the corresponding input vector and label for the query set. If episodes have training sets that are balanced $k_l = k, \forall l$, then they are called 'C way, k shot' episodes.

In this paper, a distance based meta-learner (Prototypical Networks) [4] as described below is used. The choice for this meta-learner is based on it's simplicity and robustness. The approach has been proved beneficial for few-shot classification, having a potential to outperform many sophisticated meta-learning methods [12].

Prototypical Networks: Here, a prototype of each class is constructed at the time of training. During testing, each query sample of the class is classified based on the 'nearest' Euclidean distance from the prototype. The probability that a query sample $\boldsymbol{m}^*$ belongs to class k is defined as:

$$p(n^* = k|\boldsymbol{m}^*, \mathcal{S}_s) = \frac{\exp\left(-\|g(\boldsymbol{m}^*) - \mathbf{c}_k\|_2^2\right)}{\sum_{k' \in \{1,\ldots,C\}} \exp\left(-\|g(\boldsymbol{m}^*) - \mathbf{c}_{k'}\|_2^2\right)} \tag{1}$$

where $\mathbf{c}_k$ is the 'prototype' for class k: the average of the embeddings of the training samples of class k.

3.2. Baseline approach to FSL

In this, a baseline is used for few-shot classification [11, 13]. A classifier is trained on all training classes at once, which is parameterized by the network parameters having a single output per class as shown in Fig. 1 (b). We use batch training and ResNet-18 [14] as a backbone, however also other DL model may be used. To classify new classes, the model is finetuned using a given test episode on top of g using the support set and performs classification for the query set [11, 12].

4. EXPERIMENTAL SETUP

4.1. Dataset

In this paper, we use the following datasets in RGB channels: Eurosat [15], Aerial Image Dataset (AID) [16], UCMerced Land Use [17], WHU-RS19 [18], and NWPU-RESISC45 [19]. The data is split in train (70%), validation (15%) and testing (15%) sets. Each set includes classes that are not shared with the other sets. The dataset NWPU-RESISC45 is used for testing purposes only and classes from this dataset are not included for training or validation. For evaluation, we report the mean Overall Test Accuracy (OA) and mean Cohen's kappa coefficient (κ) for each dataset [20] .

4.2. Models

4.2.1. Model A: Few-shot RS classification using meta-learning (prototypical network, as meta-learner)

The prototypical network model is trained episodically with the generated episodes using training splits of the datasets. The model is trained from scratch with the embedding function g as ResNet-18, having the input size of 256×256, with a run time as shown in Tab. 1. For episodic sampling, this paper uses variable ways/shots. The minimum of classes in any episode is set to 2 with a maximum upperbound of 50. The size of maximum number of query samples is set to 10 and support set to 500 (with maximum support size per class to 100). The model is evaluated using 600 evaluation episodes, with checkpoints recorded after every 500 steps till model reaches 50k steps. The mean Validation Accuracy (VA) is found to be 88.2694. The mean OA and mean κ (600 episodes) for different datasets is shown in Tab. 2, Model A.

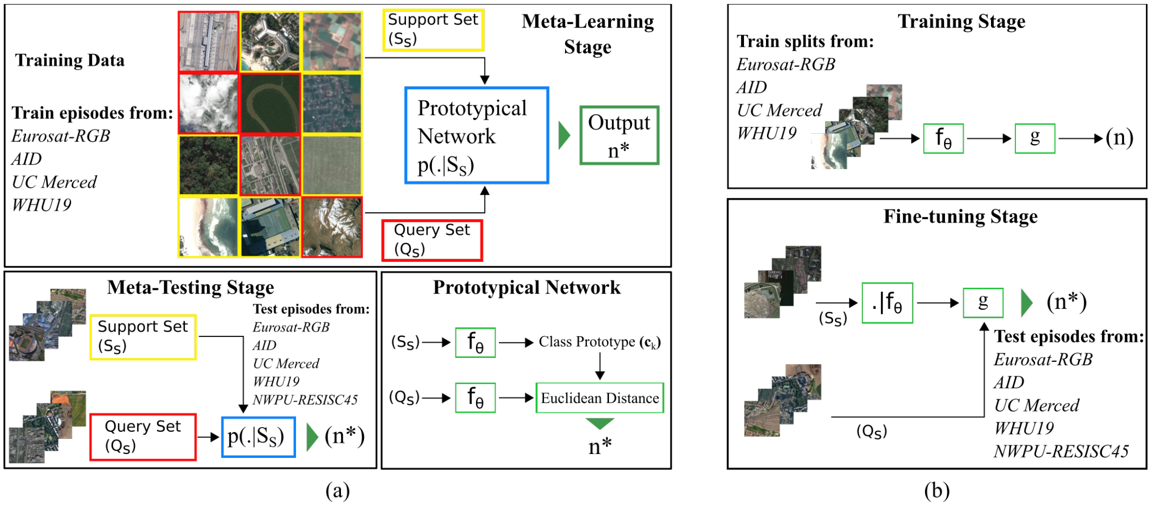

Fig. 1. (a) Meta-learning approach to FSL, and (b) baseline finetuning approach, where f_θ is the feature extraction layer parameterized by θ and g is the embedding function using the ResNet-18 architecture.

Table 1. Training Time on a Tesla V100-PCIE-32GB GPU.

Models	Model A	Model B	Model C
Training Time	45 hours	90 hours	30 hours

4.2.2. Model B: Few-shot RS classification without meta-learning (using Baseline Finetune)

Baseline Finetune model is trained from scratch with the embedding function g as ResNet-18, having the input size of 256×256 and also the batch size of 256. Finetuning with ADAM is performed, with 75 steps and learning rate of 0.01. The mean VA is found to be 92.7264. The mean OA and mean κ (600 episodes) for different datasets is shown in Tab. 2, Model B. The evaluation procedure is similar to Model A. The run time of model for training is shown in Tab. 1.

4.2.3. Model C: Few-shot RS classification using meta-learning with initialization through Baseline

The Prototypical model is initialized using the pre-trained Baseline model. The Baseline model is trained on all the datasets and then used to initialize the weight of the meta-learner. The best pre-trained Baseline model is evaluated and their parameters are used as the initialization point for the Prototypical model with the same configuration as described in Sec. 4.2.1. The mean VA is found to be 93.09583. The mean OA and mean κ (600 episodes) for different datasets is shown in Tab. 2, Model C. The run time of model for training is shown in Tab. 1.

5. RESULTS AND DISCUSSION

Tab. 2 shows that Model C outperforms both the Model A and Model B in terms of their OA and κ for the test classes of all datasets and with the smallest run time. In addition to this, it could also be seen that Model A takes half the training time as compared to Model B. However, Model B performs better for an unseen dataset NWPU-RESISC45 having the potential to be used as a better initializer. These results imply that the meta-learning approach can be beneficial for few-shot RS classification, if we carefully choose the model parameters or use a proper initialization technique. Using the meta-learning approach with proper initialization also reduces the overall training (compute) time of the model which becomes essential with large datasets.

In addition, adapting the meta-dataset pipeline for RS images, helps to give diversity to the training and test data which is important to study how well our model is in terms of generalization, even in presence of limited number of training samples. Moreover, the lower values of κ for the NWPU-RESISC45 and AID datasets for all the models suggests that there is a higher class imbalance in these datasets which needs to be accounted. Apart from this, the Baseline approach is also having a potential, however in future, carefully choosing the backbone and model parameters can result in a better training time, thereby giving this approach a better chance.

6. CONCLUSIONS

We performed few-shot RS classification using meta-learning (prototypical network as a meta-learner) using RS datasets (Eurosat, Aerial Image Dataset, UCMerced Land Use, WHU-RS19, and NWPU-RESISC45) as leveraged from the adapted meta-dataset pipeline. For comparison purposes, we also perfomed few-shot RS classification using a Baseline model without meta-learning. We could see that even in the presence of a limited amount of training samples, a comparable generalization performance was acheived using meta-learning approaches in Few Shot Learning. Also, with a proper initialization, the approach helps to generalize on the unseen data, with a significantly shorter training time.

Table 2. Few-shot RS classification mean Overall Test Accuracies OA in percentage and mean κ computed on 600 episodes. (The percentage in the Dataset column shows the reserved data split for testing purposes)

Test classes for Datasets	**Model A**		**Model B**		**Model C**	
	OA	κ	OA	κ	OA	κ
EuroSAT (15%)	98.4750 +/- 0.2863	0.972333	97.4750, +/- 0.3630	0.946167	99.1833, +/- 0.1833	0.988167
AID (15%)	78.3736 +/- 0.8904	0.668083	82.9958, +/- 0.7779	0.722250	85.8972, +/- 0.7071	0.781139
UC Merced (15%)	88.7305, +/- 0.7843	0.814000	95.1583, +/- 0.4619	0.920333	95.2000, +/- 0.4841	0.929583
WHU19 (15%)	99.2306, +/- 0.2274	0.987167	98.6278, +/- 0.2529	0.982000	99.7694, +/- 0.1032	0.994333
NWPU-RESISC45 (100%)	69.0711, +/- 1.0950	0.596833	81.8329, +/- 0.8740	0.747792	79.9027, +/- 0.8811	0.738003

7. REFERENCES

[1] A. Stoian, V. Poulain, J. Inglada, V. Poughon, and D. Derksen, "Land Cover Maps Production with High Resolution Satellite Image Time Series and CNNs: Adaptations and Limits for Operational Systems," *Remote Sensing*, vol. 11, no. 17, 2019.

[2] S. Thrun and L. Pratt, Eds., *Learning to Learn*, Kluwer Academic Publishers, 1998.

[3] Li Fei-Fei, R. Fergus, and P. Perona, "One-Shot Learning of Object Categories," *IEEE TPAMI*, vol. 28, no. 4, pp. 594–611, 2006.

[4] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-Shot Learning," in *Advances in Neural Information Processing Systems*, 2017.

[5] L. Zintgraf, K. Shiarlis, V. Kurin, K. Hofmann, and S. Whiteson, "Fast Context Adaptation via Meta-Learning," in *Proceedings of the 36th ICML*, 2019.

[6] J. Gordon, J.and Bronskill, S. Nowozin, M. Bauer, and R. E. Turner, "Meta-learning probabilistic inference for prediction," in *Proceedings of the 7th ICLR*, 2019.

[7] M. Zhai, H. Liu, and F. Sun, "Lifelong Learning for Scene Recognition in Remote Sensing Images," *IEEE GRSL*, vol. 16, no. 9, pp. 1472–1476, 2019.

[8] D. A. Alajaji and H. Alhichri, "Few Shot Scene Classification in Remote Sensing using Meta-Agnostic Machine," in *Proceedings of the 6th CDMA*, 2020, pp. 77–80.

[9] M. Wang, S.and Rußwurm, M. Körner, and D. B. Lobell, "Meta-Learning For Few-Shot Time Series Classification," in *IEEE IGARSS*, 2020, pp. 7041–7044.

[10] Z. Yuan, W. Huang, L. Li, and X. Luo, "Few-Shot Scene Classification With Multi-Attention Deepemd Network in Remote Sensing," *IEEE Access*, vol. 9, pp. 19891–19901, 2021.

[11] C. Wei-Yu, L. Yen-Cheng, K. Zsolt, F. Wang Yu-Chiang, and H. Jia-Bin, "A Closer Look at Few-shot Classification," 2020.

[12] E. Triantafillou, T. Zhu, V. Dumoulin, P. Lamblin, U. Evci, K. Xu, R. Goroshin, C. Gelada, P. A. Swersky, K. Manzagol, and Larochelle H., "Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples," 2019.

[13] G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, "A baseline for few-shot image classification," in *International Conference on Learning Representations*, 2020.

[14] K. He, R. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015.

[15] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification," in *IEEE IGARSS*, 2018, pp. 204–207.

[16] G. S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A Benchmark Data set for Performance Evaluation of Aerial Scene Classification," *IEEE TGRS*, 2017.

[17] Yi Yang and Shawn Newsam, "Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification," in *Proceedings of the ACM GIS*, 2010.

[18] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Maitre, and H. Sun, "Structural High-Resolution Satellite Image Indexing," *ISPRS*, 2010.

[19] G. Cheng, J. Han, and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," 2017.

[20] Margherita Grandini, Enrico Bagli, and Giorgio Visani, "Metrics for multi-class classification: an overview," 2020.