Model-Agnostic Meta-Learning Networks for Remote Sensing Scene Classification

Adam DiChiara[1] **Vincent Filardi**[1] **Quincy Hershey**[1] **Alexander Moore**[1] **Scott Tang**[1]

Abstract

Our work addresses generalizability problems in the field of remote sensing scene classification (RSSC) using a novel approach of model-agnostic meta-learning (MAML) and task simulation. The remote sensing scene classification problem is inherently challenging due to data variance and data quality factors [6,10,15,16], as well as challenges with data intensity and transferability to other target data sources [6,15]. We present a novel MAML approach to this problem, which utilizes image transforms to simulate dataset tasks as a way of improving learning of semantic content within scene classes. The goal of this work is to produce a few-shot model that is more generalizable and requires less training data than the currently dominant models, while also achieving similar or better success against benchmarks. We implemented a task-specific MAML first order approximation model (with and without L2 regularization) and a generalized MAML CNN in the n-way k-shot setting, as well as CNN baseline models (with and without data augmentation). We evaluate transferability by applying these to a secondary target RSSC dataset. Our MAML results did not exceed baseline Augmented CNN performance, which we attribute in part to task-similarity of our augmentations and class-similarity between datasets.

1. Introduction

Remote sensing imaging refers to visual images of earth objects and scenes taken from remote aircraft or satellites (Fig 1). Major applications involve earth sciences, mapping, urban planning, environmental monitoring, natural hazard detection, surveillance, and reconnaissance [4,14]. Remote sensing image classification can be divided into three main task areas: pixel-based classification (i.e., semantic segmentation), object-level classification (i.e., object recognition), and scene-level classification. Scene classification involves correctly labelling cropped portions of large-scale images into semantic categories (e.g., “freeway”, “river”, “commercial area”), and has become an increasingly active research area for deep learning methods due to several inherent challenges with this task [6,11,15,16]. These include high within-class diversity (e.g., schools often appear visually different in style or shape), high inter-class similarity (e.g., both “bridge” and “overpass” scene classes contain many of the same objects), a large variance in the scales of objects and scenes (i.e., images are often taken at various altitudes), and the co-presence of multiple objects within scenes [6,14,16]. Additionally, image quality issues and atmospheric effects can affect feature learning [6,15].

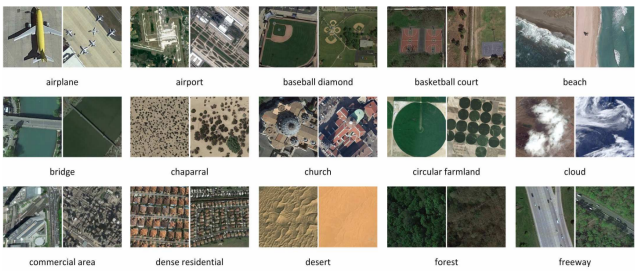

Figure 1: Examples of remote sensing scene classes.

While the best performing RSSC models to date, primarily convolutional neural networks (CNNs), have

[1] Worcester Polytechnic Institute. Correspondence to: Adam DiChiara <ajdichiara@wpi.edu>
https://github.com/moorea1/CS541_Group

CS/DS541 Deep Learning Final Project, Worcester Polytechnic Institute.

demonstrated success against available benchmarks, the data variance issues are not yet considered solved problems. Compounding this challenge is the lack of available training data. As a result, there has been little success in transferring high performing models to other RSSC target data sources [6,15].

This project focuses on the specific problem of generalizability in scene classification. In exploring the applications of rapidly adaptable frameworks in novel architectures, we hope to add further insight into what methods may be effective in best approaching the issues of generalizability and data intensity in remote scene classification. We examine and implement several competing network architectures and training algorithms in an attempt to develop well performing solutions that could be utilized in different real world applications.

1.1 Research contributions

In comparison to the currently dominant RSSC approaches, we recognize the need for a robust, generalizable, less data hungry strategy that can demonstrate high performance against existing benchmarks without suffering from the limitations described previously. A review of the recent surveys of the field suggests that our MAML implementation with augmentation-based task simulation is an approach that is both novel and worthy of investigation [4,6,9,15].

2. Related Work

The best performing CNN-based deep learning RSSC strategies have been shown to be prone to overfitting for scene classification and problematic when applying to RSSC target domains that differ from the training source [6,9,10,11,13,16]. Research has shown these models are vulnerable to image quality and resolution factors, atmospheric effects, scale variance, and other image variations [6,11,15]. They also require large sources of annotated training data for each scene class which is limited and prohibitive to acquire in the RSSC domain [3,6,9,10,11,16]. Given the lack of available labeled training data, attention in the field is beginning to turn to few-shot/zero-shot solutions; however, this is an extremely recent development and requires further exploration [6,9].

3. Proposed Method

The second-order optimization scheme MAML is a meta-learning framework in the few-shot setting. The algorithm aims to prime the model's weights for fast adaptation within a few steps of updates and a small amount of data. We implemented the first order variant of MAML which reduces computational complexity by omitting second derivatives required during optimization in the general-form MAML implementation [7]. The strength of MAML comes from the relationship between the shared weight initialization of each task on the inner loop and the update step with respect to the initialization on the inner loop. The First-Order MAML approximation, FOMAML, severs this relationship by taking the gradient with respect to individual task parameters instead of the initial weights [7]. To force the relationship between the initial parameters, other papers have imposed penalties on the inner loop's gradient update with high success such as implicit MAML [11]. Similarly, using this as justification we impose an L2 penalty in the inner loop of first order MAML to stay closer to the initial weights of the task loop.

We induced the novel task-augmentation as our task distribution using combinations of image transforms: flips, rotation, cropping, perspective, gaussian noising, and brightness/contrast. Importantly these augmentations would be expected to be encountered in remote sensing data and are relevant to the challenges of scene classification. For instance, cropping induces scale variance and noising may simulate image quality disparities. In our implementation, a given task consists of between zero and t=7 max augmentations to be applied to a batch of images. The augmentation parameters for each transform were set randomly within a preset range for each batch associated with that task. We defined our task structure as a full-factorial binary matrix of augmentation combinations, having length 2^t, to obtain 128 total tasks. By introducing these augmentations at training time, our intent was to create a model which emphasizes learning the semantic content of images, in hopes that at test-time on the same tasks with unseen data, classification performance will be improved. Such a method should theoretically lower the amount of data needed for generalization.

Under the framework of a MAML optimization scheme, as described in [8], our expectations were that the model would not overfit to any of the transformations and would learn more robust representations of the scene data. In addition, the framework's few-shot setting will allow us to work with smaller amounts of data. Both traits are highly desirable in RSSC, as discussed previously.

Initial structures we employed examined the application of auto-encoders as feature-learning models but found they were unable to compete empirically with the results generated by a CNN. We had theorized that by learning a low-dimensional, information-dense bottleneck layer we might improve performance of a downstream supervised model. Following the inability of auto-encoders to deliver performance on par with a CNN, our focus shifted to delivering rapidly adaptable results through the use of MAML trained CNN architectures.

4. Experiment

The datasets used for model training and testing and the evaluation methods used are described here. Torchmeta was used for gradient management and learn2learn for the few-shot setup [2,7].

4.1 Data Sources

This project makes use of the publicly available NWPU-RESISC45 dataset (2016) (RESISC), which contains 31,500 total remote sensing images comprising 45 scene classes and currently represents the largest dataset for RSSC benchmarking [5,16]. We chose this dataset because it presents many generalizability challenges, including high variance in image quality, resolution, object pose, and object scale. It also contains high within-class diversity and high between-class similarity, which are two of the most significant challenges for RSSC [5,6,14]. To evaluate transferability to new remote sensing images, we also apply our models to the UC-Merced Land Use dataset (UCMerced) which includes 2,100 images evenly distributed among 21 classes for transfer learning analysis following training on RESISC. To better create a distinct comparison, overlapping classes were removed from RESISC, resulting in two completely distinct sets of labels. After this reduction, the RESISC dataset contained 18,200 images across 26 classes.

Table 1: Dataset Attributes.

Dataset	Year	Source	Scene Classes	Images per Class	Total Images	Image Size	Type
RESISC	2016	Google Earth	45	700	31500	256x256	RGB
UCMerced	2010	USGS National Map Urban Area Imagery	21	100	2100	256x256	RGB

4.2 Experiment Structure

We used the following roadmap for models and methods:

1. Train competing infrastructures on the RESISC dataset using stratified splits with 20% training data of which 20% utilized as a validation set with individually tuned hyperparameters (See GitHub[1]).
2. To create a more distinct pairing of training and transfer learning datasets, 19 of 46 overlapping classes were removed from the RESISC dataset.
3. Initially train a 5-layer CNN with layers increasing in output by a factor of 2 starting from 32, utilizing ReLU activation and batch norm layers to establish baseline performance (CNN).
4. Introduce data augmentations based on a single task structure to the (3) paradigm and record testing performance (CNN Augmented).
5. Introduce the MAML framework and task structure to the (3) paradigm, using 32-way 1 shot, 32-task batch size, and record testing performance across all data augmentation tasks. (MAML)
6. Introduce first-order approximation MAML framework, using 1-task batch size. (FOMAML)
7. Introduce first-order approximation MAML framework, using 32-task batch size and L2 regularization (FOMAML+L2)
8. Port models to the UCMerced dataset by exchanging the 26 class classifier layer for a 21 class classifier suited for the new dataset.
9. Allow each model 3 epochs of training over a stratified 20% training split (420 images) before testing.

10. Measure transferability on a new unseen similar dataset with the best performing model. To do this we take an average of 20 training runs on UC dataset for 3 epochs each on the pretrained model as in (8). We report the average test accuracy of those 20 training runs.

4.3 Evaluation Metrics

Model performance is reported as overall classification accuracy (i.e., total correctly classified scene images / total images) as well as overall "top 3" accuracy (i.e., percent of scene images for which true classification was within highest three predicted probabilities). We present accuracy results obtained from two baseline models (CNN and CNN Augmented) and three MAML versions (MAML, FOMAML, FOMAML+L2) for both the RESISC and UCMerced datasets.

5. Results

The Augmented CNN outperformed all other models within both the RESISC dataset and the UCMerced dataset at 70% and 67% respective accuracies. The performance of CNN Augmented was followed closely by the FOMAML+L2 at 64% on the RESISC dataset and 64% on the UCMerced dataset. The standard CNN and MAML both did not compete on par with the others (Table 2). The CNN rapidly began to overfit and quickly reached its peak validation performance within a period of 20 epochs while the generalized MAML possibly struggled to find optimized hyperparameters. A common shortcoming of MAML [1].

Table 2: Results of model experiments.

Best Results	CNN	CNN Augmented	FOMAML	FOMAML+ L2	MAML
RESISC Accuracy	34.70%	**70.60**%	60.26%	63.99%	52.47%
RESISC Top 3 Accuracy	55.43%	**89.50**%	84.19%	85.28%	75.09%
UCMerced Accuracy	45.05%	**67.12**%	62.19%	64.53%	26.57%
UCMerced Top 3 Accuracy	69.35%	**88.34**%	85.10%	86.60%	47.32%

6. Discussion

The results of this research are summarized by the classification accuracy of each proposed model detailed (Table 2). These results show that meta-learning struggled to compete with the heuristic convolutional neural network classifier with task structure. Introducing task structure did not increase performance in all models besides the Augmented CNN. Similarly in testing, the n-way k-shot setting degraded performance across the board.

It is possible that there are changes that would improve the performance of the meta-learning classifiers. Since we designed the tasks for the meta-learning algorithm to be compositions of augmentations, there are likely superior choices of augmentations. In a small test we found performance across all MAML variants improved with fixed rotations per batch as one of the augmentations per task but did not outperform the CNN with the random rotation included in the task structure. The augmentations we chose are ones that would possibly occur in remote sensing images, but using augmentations in the typical supervised learning capacity is demonstrated to be superior in this example. It is also the case that the variation between tasks created through data augmentation may not be substantial enough to be well suited to the MAML algorithm which seems to perform better in settings where the distribution between tasks is more multi-modal. In addition, further hyperparameter and architecture tuning for the meta-learning model could improve performance as we found the supervised accuracy of the meta-learner to change dramatically with the ratio of inner to outer learning rates.

7. Conclusions and Future Work

The Augmented CNN architecture narrowly outperformed FOMAML+L2 and outperformed the CNN and MAML by a larger margin on both the RESISC and UCMerced datasets. The results may be attributable to task selection and tuning, hyperparameter selection, similarities between both datasets or simply the strength of the Augmented CNN. Future work may focus on further tuning the network structures as well as redefining the task structure to include more modal diversity between datasets. Additionally, choosing a more dissimilar

dataset for transfer learning may improve relative performance of the MAML algorithm versus the Augmented CNN.

8. References

[1] Antoniou, A., Edwards, H., & Storkey, A. (2018). How to train your MAML. *arXiv preprint*. arXiv:1810.09502

[2] Arnold, S. M., Mahajan, P., Datta, D., Bunner, I., & Zarkias, K. S. (2020). learn2learn: A Library for Meta-Learning Research. *arXiv preprint*. arXiv:2008.12284

[3] Ball, J. E., Anderson, D. T., & Chan, C. S. (2018). Special section guest editorial: feature and deep learning in remote sensing applications. *Journal of Applied Remote Sensing, 11(4)*, 042601. https://doi.org/10.1117/1.JRS.11.042601

[4] Ball, J. E., Anderson, D. T., & Chan Sr., C. S. (2017). Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing, 11(4)*, 042609. https://doi.org/10.1117/1.JRS.11.042609

[5] Cheng, G., Han, J., & Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE, 105(10)*, 1865-1883. https://doi.org/10.1109/JPROC.2017.2675998

[6] Cheng, G., Xie, X., Han, J., Guo, L., & Xia, G. S. (2020). Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3735-3756. https://doi.org/10.1109/JSTARS.2020.3005403

[7] Deleu, T., Würfl, T., Samiei, M., Cohen, J. P., & Bengio, Y. (2019). Torchmeta: A Meta-Learning library for PyTorch. *arXiv preprint*. arXiv:1909.06576.

[8] Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning* (pp. 1126-1135). PMLR. https://arxiv.org/abs/1703.03400v3

[9] Li, H., Cui, Z., Zhu, Z., Chen, L., Zhu, J., Huang, H., & Tao, C. (2020). RS-MetaNet: Deep meta metric learning for few-shot remote sensing scene classification. In *IEEE Transactions on Geoscience and Remote Sensing (Early Access)*, 1-12. https://doi.org/10.1109/TGRS.2020.3027387

[10] Li, Y., Zhang, Y., & Zhu, Z. (2019, July). Learning deep networks under noisy labels for remote sensing image scene classification. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3025-3028). IEEE. https://doi.org/10.1109/IGARSS.2019.8900497

[11] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing, 152*, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

[12] Rajeswaran, A., Finn, C., Kakade, S., & Levine, S. (2019). Meta-learning with implicit gradients. *arXiv preprint*. arXiv:1909.04630

[13] Song, S., Yu, H., Miao, Z., Zhang, Q., Lin, Y., & Wang, S. (2019). Domain adaptation for convolutional neural networks-based remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, *16*(8), 1324-1328. https://doi.org/10.1109/LGRS.2019.2896411

[14] Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., ... & Lu, X. (2017). AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 55(7)*, 3965-3981. https://doi.org/10.1109/TGRS.2017.2685945

[15] Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine, 4(2)*, 22-40. https://doi.org/10.1109/MGRS.2016.2540798

[16] Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine, 5(4)*, 8-36. https://doi.org/10.1109/MGRS.2017.2762307