TIML: Task-Informed Meta-Learning for crop type mapping

Gabriel Tseng[1,2], **Hannah Kerner**[1,3], **David Rolnick**[2]

[1]NASA Harvest
[2]McGill University and Mila – Quebec AI Institute
[3]University of Maryland, College Park

Abstract

Labeled datasets for agriculture are extremely spatially imbalanced. When developing algorithms for data-sparse regions, a previously explored approach is to use transfer learning from data-rich regions. While standard transfer learning approaches typically leverage only direct inputs and outputs, geospatial imagery and agricultural data is rich in metadata that can inform transfer learning algorithms, such as the spatial coordinates of data-points. We build on previous work exploring use of meta-learning to crop type mapping in data-sparse regions and introduce task-informed meta-learning (TIML), an augmentation to model-agnostic meta-learning which takes advantage of this metadata. We apply TIML to the CropHarvest dataset, a global dataset of agricultural class labels paired with remote sensing data. In addition, we introduce the concept of forgetfulness when training meta-learning models on many similar tasks to mitigate memorization of training tasks. We find that TIML significantly improves average performance across the CropHarvest evaluation tasks compared to a range of benchmark models, measured using AUC ROC and F1 scores.

Introduction

The global food system both drives the climate crisis—contributing to 26% of global greenhouse gas emissions (Ritchie 2019)—and is vulnerable to it, with every 1% increase in temperature lowering yields by 2.28% (Ozdemir 2021). Crop maps, which spatially identify where crops are being grown, provide vital information for mitigating and adapting to the effects of climate change, for example by assessing food security, more rapidly responding to food crises, and increasing productive land without sacrificing carbon sinks such as forests.

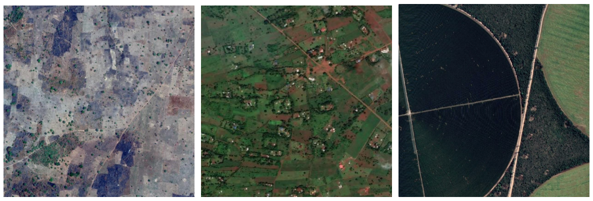
Certain parts of the world collect plentiful field-level agricultural data, but many regions are extremely data-sparse. We investigate transfer learning from data-rich areas to improve performance in data-sparse areas, building on the meta-learning approach described in Rußwurm et al. (2020) and Tseng et al. (2021a,b). Meta-learning aims to learn a model - from a number of training tasks - which can quickly learn a new task from a small amount of new data. To construct these tasks, we split a dataset with global coverage into multiple geographically and semantically defined tasks. A meta-learning model is trained on a subset of these tasks (using model-agnostic meta-learning, or MAML (Finn, Abbeel, and Levine 2017)), before being evaluated on the remaining tasks. In this work, we consider the meta-information captured in each task (such as the geography being covered by a task, or the specific classes being classified) and how this might be passed to the meta-learning model to improve performance.

We summarize the main contributions of this paper below:

- We introduce Task-Informed Meta-Learning (TIML), an algorithm designed to augment model-agnostic meta-learning with task-metadata.
- We apply TIML to crop type classification and demonstrate it works well across a wide range of agroecologies and crops, outperforming other methods on both AUC ROC and F1 score.
- In particular, we highlight TIML's ability to learn from very few positive labels and to perform well on tasks where other transfer-learned models do poorly.

Related Work

In applications of machine learning to remote sensing data, there have been numerous efforts to learn from data-rich areas and transfer the resulting model to data-sparse regions or less well represented classes, including meta-learning (Rußwurm et al. 2020; Tseng et al. 2021a,b), transfer learning (Wang et al. 2018; Tong et al. 2020) and multitask learning (Kerner et al. 2020). However, these approaches often fail to capture important meta-data and expert knowledge about the data-sparse tasks of focus, such as their geographic location relative to the pre-training data or the crop types being classified. In this work, we consider the additional metadata inherent to agricultural classification tasks, such as the spatial relations between tasks, and how this can inform the model's predictions. This builds on previous work that has investigated shifting the distribution of the meta-model with respect to the task being considered (Vuorio et al. 2019; Triantafillou et al. 2021). However, we consider the special case where there is task-specific metadata that stays static for each task and can be used to condition the model.



(a) Togo (b) Kenya (c) Brazil

Figure 1: Example 1km × 1km satellite images of the evaluation regions, demonstrating the variety in field sizes and agro-ecologies being evaluated. (Images were obtained from Google Earth Pro basemaps comprised primarily of high resolution Maxar images, and are reproduced with permission from (Tseng et al. 2021b))

Data

We use the CropHarvest dataset (Tseng et al. 2021b) to train the model. This dataset consists of 90,480 datapoints with the associated satellite pixel time-series for each point. Of these datapoints, 30,899 (34.2%) contain multi-class agricultural labels; the remaining datapoints contain binary "crop" or "non-crop" labels. Each datapoint is accompanied by a pixel time-series from 4 remote sensing products: Sentinel-2 L1C optical observations, Sentinel 1 synethic aperture radar, ERA5 climatology data, and topography from a Digital Elevation Model (DEM), representing 1 year of data at monthly timesteps.

The CropHarvest dataset is additionally accompanied by 3 evaluation tasks which test the ability of a pre-trained model to learn from a small number of in-distribution datapoints in a variety of agroecologies. Satellite imagery from each test task is shown in Figure 1, highlighting the variety of agroecologies covered in the test tasks. We describe each task and the accompanying training data below:

Togo crop vs. non-crop: Classifying pixels containing crops from those which do not, in Togo. The training set consists of 1,319 datapoints and the test set consists of 306 datapoints - 106 (35%) positive and 200 (65%) negative - sampled from random points within the country.

The two other evaluation tasks consist of classifying a specific crop. Thus, "rest" below refers to all other crop and non-crop classes. For both tasks, entire polygons delineating a field (as opposed to single pixels) were collected, allowing evaluation across the polygons. However, during training only the polygon centroids were used.

Kenya maize vs. rest: The training set consists of 1,345 imbalanced samples (266 positive and 1,079 negative samples). The test set consisted of 45 polygons ultimately yielded 575 (64%) positive and 323 (36%) negative pixels.

Brazil coffee vs. rest: The training set consists of 794 imbalanced samples (21 positive and 773 negative samples). The test set consisted of 66 polygons, which yielded 174,026 (25%) positive and 508,533 (75 %) negative pixels.

Methods

We focus on a specific meta-learning method, Model-Agnostic Meta-Learning (MAML) (Finn, Abbeel, and Levine 2017). MAML learns set of model weights θ which are close to the optimal weights for a variety of different task, allowing the optimal weights for a specific task to be reached with little data and/or few gradient steps. These initial weights θ are updated by finetuning them on a training task (inner loop training), yielding updated weights θ'. A gradient for θ is then computed with respect to the loss of the updated model, $L_{\theta'}$ (requiring Hessian vector-products). This gradient is then used to update θ (outer loop training).

MAML with the CropHarvest dataset

As with the CropHarvest benchmarks, we defined tasks spatially using bounding boxes for countries drawn by Natural Earth (Patterson and Kelso). Tasks consist of binary classification of pixels as either crop vs. non-crop or a specific crop-type vs. rest. This yielded 525 tasks, which were split into training and validation tasks, from which the model could learn. The 3 evaluation tasks described previously were withheld from the initial training. Then for each evaluation task, we fine-tuned the model on that task's training data before evaluating the model on that task's test data.

Task-Informed Meta-Learning

We build on the original model-agnostic meta-learning (Finn, Abbeel, and Levine 2017) algorithm, considering the case where there is additional task-specific information that could inform the model, such as the spatial relationships between tasks. Information such as the spatial coordinates of a task (which we represented as the central coordinates of the task-country) remains static for all datapoints in the task, so is not useful to differentiate positive and negative instances. However, it may be useful to condition the model prior to inner loop training.

Algorithm 1: Task-Informed Meta-Learning

Require: $p(\mathcal{T})$: Distribution over tasks
Require: α, β: step size hyperparameters
randomly initialize meta model θ_m, task encoder θ_e
while not done **do**
    Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$ with task information t_i
    for all $\mathcal{T}_i, t_i$ **do**
        Generate task embeddings $\mu_i = f(t_i; \theta_e)$
        Evaluate $\nabla_{\theta_m} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_m}, \mu_i)$ with respect to K examples
        Compute adapted meta parameters with gradient descent: $\theta^{'}_{m_i} \leftarrow \theta_m - \alpha \nabla_{\theta_m} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_m}, \mu_i)$
    Update $\theta_m \leftarrow \theta_m - \beta \nabla_{\theta_m} \Sigma_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_{m_i}}, \mu_i)$
    Update $\theta_e \leftarrow \theta_m - \beta \nabla_{\theta_e} \Sigma_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_{m_i}}, \mu_i)$

We introduce Task-Informed Meta-Learning (TIML) (Algorithm 1), which modulates the hidden vectors in the meta-model based on embeddings calculated using task information. We use feature-wise linear modulation, or FiLM (Perez et al. 2018)), to modulate the hidden vectors. These embeddings are updated in the outer loop of the MAML training procedure.

For the CropHarvest dataset, task information is encoded in a 13-dimensional vector. Three dimensions are used to encode spatial information, consisting of latitude and longitude transformed to $[cos(\mathrm{lat}) \times cos(\mathrm{lon}), cos(\mathrm{lat}) \times sin(\mathrm{lon}), sin(\mathrm{lat})]$. This transforms the spatial information from spherical to cartesian coordinates, ensuring transformed values at the extreme longitudes are close to each other. The remaining 10 dimensions are used to communicate the type of task the model is being asked to learn. This consists of a one-hot encoding of crop categories from the UN Food and Agriculture Organization (FAO) indicative crop classification (fao 2020), with an additional class for non-crop. For crop vs. non-crop tasks, all crop type categories contain the value $\frac{1}{n}$, for n crop type categories.

We used a task encoder to learn the embeddings. This encoder consists of linear blocks, where each block contains a linear layer with a GeLU activation (Hendrycks and Gimpel 2016) and Dropout (Srivastava et al. 2014). The task information was encoded into a hidden-task vector. Independent blocks were then used to generate an embedding for each hidden vector in the classifier to be modulated. Due to the limited number of country-FAO category combinations, we additionally applied Gaussian noise to the task information when training the model.

Forgetful Meta-Learning

Since the CropHarvest dataset is global and tasks are geographically defined, the training data for the meta-learning model are divided into many discrete tasks (525). However, these tasks are not semantically or geographically well distributed. A significant fraction of the tasks are crop vs. non-crop tasks, reflecting the large number of binary crop vs. non-crop datapoints in the dataset (65.8% of all instances only have crop vs. non-crop labels). Crop-type tasks are also geographically concentrated in countries in which crop-type labels (as opposed to crop vs. non-crop labels) were collected. Finally, multiple tasks can contain similar datapoints, as positive instances for one task could be part of the negative class for a different task in the same geography. We therefore hypothesize that the model can memorize many similar tasks, to the detriment of its ability to learn more difficult or rarer tasks, thus hurting generalization performance for the fine-tuning tasks.

Although complex meta-learning methods exist designed to optimize for performance on highly challenging tasks (Jamal and Qi 2019; Collins, Mokhtari, and Shakkottai 2020), we take advantage of the large number of similar tasks in the CropHarvest dataset to introduce a very simple method to prevent memorization of certain tasks: removing training tasks the model has memorized. We define memorization as achieving an average AUC ROC of more than 0.95 on the training data over 20 consecutive epochs. Since during training each task is sampled to have a balanced number of positive and negative examples, this AUC ROC is calculated on a balanced dataset. We call this method "forgetfulness."

Experiments

We evaluated TIML by training it on the CropHarvest dataset and fine-tuning it on the evaluation tasks, as was done for the benchmark results released with the dataset in Tseng et al. (2021b). MAML (and by extension, TIML) can be applied using any neural network architecture. We used the same base classifier, a 1-layer LSTM model followed by a linear classifier, and same hyperparameters as in the CropHarvest benchmarks.

Ablations

To understand the effects of different components TIML on overall model performance, we run 3 ablations of TIML:

- **No forgetfulness**: A TIML model trained without forgetfulness; no tasks are removed in the training loop
- **No encoder**: A TIML model with no encoder. The task information is instead appended to every raw input timestep, and passed directly to the classifier.
- **No task information or encoder**: No task information passed to the model at all. This model is effectively a normal MAML model, trained with forgetfulness.

Baselines

We compared the TIML architecture to 4 baselines. As with TIML, we finetuned these models on each benchmark task's training data and then evaluated them on the task's test data:

- **MAML**: A model-agnostic meta-learning classifier without the task information.
- **Crop pre-training**: A classifier pre-trained to classify all data as crop or non-crop.
- **No pre-training**: A randomly initialized classifier, which is not pre-trained on the global CropHarvest dataset but instead is trained directly on the test task training data.

For these LSTM-based classifiers (as for TIML and its ablations), we fine-tuned the models on the test tasks for 250 gradient steps with batches containing 10 positive and 10 negative examples (as in Tseng et al. (2021b)).

In addition, we trained a **Random Forest** baseline, implemented using scikit-learn (Pedregosa et al. 2011) with the default hyperparameters.

Results & Discussion

Model results for TIML, its ablations and all baseline models are shown in Table 1. We report the AUC ROC score and the F1 score calculated using a threshold of 0.5. Overall, TIML is the best performing algorithm on the CropHarvest dataset, achieving the highest F1 scores and AUC ROC scores when averaged across all tasks. TIML is consistently among the best performing algorithms when considering specific tasks, and in particular is the only transfer-learning model able to outperform a randomly-initialized model in the Brazil task when measured by the F1 score. We note the very small number of positive datapoints (26) in the Brazil task, and the comparative weakness of other transfer-learning models on this task.

Effects of transfer learning Standard transfer learning from the global dataset is not guaranteed to confer advantages to the model. In Kenya, all transfer learning models see significant increases in performance irrespective of the

Table 1: Results for the evaluation tasks. All results are averaged from 10 runs and reported with the accompanying standard error. We report the area under the receiver operating characteristic curve (AUC ROC) and the F1 score using a threshold of 0.5 to classify a prediction as the positive or negative class. We highlight the **first** and **second** best metrics for each task. TIML achieves the highest F1 score of any model on the Brazil task and the best AUC ROC and F1 scores when averaged across the 3 tasks. We highlight the improvement of TIML relative to other transfer-learning models, suggesting it is better able to model the diversity in tasks present in the CropHarvest dataset.

	Model	Kenya	Brazil	Togo	Mean
AUC ROC	Random Forest	0.578 ± 0.006	0.941 ± 0.004	0.892 ± 0.001	0.803
	No pre-training	0.329 ± 0.011	0.898 ± 0.010	0.861 ± 0.002	0.700
	Crop pre-training	0.694 ± 0.001	0.820 ± 0.002	$\mathbf{0.894 \pm 0.000}$	0.801
	MAML	$\mathbf{0.729 \pm 0.001}$	0.831 ± 0.005	0.878 ± 0.001	0.843
	TIML	$\mathbf{0.794 \pm 0.003}$	$\mathbf{0.988 \pm 0.001}$	0.890 ± 0.000	**0.890**
	no forgetfulness	0.779 ± 0.003	0.877 ± 0.003	0.893 ± 0.001	0.850
	no encoder	0.712 ± 0.001	$\mathbf{0.977 \pm 0.002}$	$\mathbf{0.895 \pm 0.000}$	**0.862**
	no task info or encoder	0.690 ± 0.001	0.977 ± 0.002	0.876 ± 0.001	0.848
F1 score	Random Forest	0.559 ± 0.003	0.000 ± 0.000	$\mathbf{0.756 \pm 0.002}$	0.441
	No pre-training	0.782 ± 0.000	$\mathbf{0.764 \pm 0.012}$	0.720 ± 0.005	**0.734**
	Crop pre-training	0.819 ± 0.001	0.619 ± 0.005	0.713 ± 0.002	0.613
	MAML	0.828 ± 0.001	0.496 ± 0.001	0.662 ± 0.001	0.652
	TIML	$\mathbf{0.838 \pm 0.000}$	$\mathbf{0.835 \pm 0.012}$	0.732 ± 0.002	**0.802**
	no forgetfulness	$\mathbf{0.840 \pm 0.000}$	0.537 ± 0.002	$\mathbf{0.764 \pm 0.002}$	0.724
	no encoder	$\mathbf{0.840 \pm 0.000}$	0.473 ± 0.002	0.691 ± 0.001	0.691
	no task info or encoder	0.837 ± 0.001	0.473 ± 0.001	0.645 ± 0.002	0.652

transfer learning method. However, this is not the case in Brazil where first training using MAML or crop pre-training penalizes the model performance compared to an LSTM initialized with random weights. We hypothesize this may be due to the difference in distribution of the Brazil-coffee tasks relative to the other tasks the models are trained on. TIML is the only model to see significant general improvements in performance compared to the randomly initialized model, suggesting conditioning the model with prior, domain-specific information about the tasks can help to model the diversity of samples in the CropHarvest dataset.

Forgetfulness We compare the effects of training TIML with and without forgetfulness, and find that forgetfully significantly boosts performance in Brazil without significantly impacting performance on other tasks. This combination (TIML with forgetfulness) ultimately yields significantly higher mean F1 and AUC ROC scores when measured across all tasks. However, training TIML forgetfully without the task information (TIML with no task information or encoder) yields comparable results to the baseline MAML model trained without forgetfulness. We therefore hypothesize that task information provides useful context around which tasks are being kept and forgotten during training, allowing TIML to learn from more difficult tasks in the "forgetful" regime without forgetting easier tasks it has already learned.

Effect of task information Including task information in the model improves performance, both when it is concatenated to the input data and when it is passed to the model through TIML. However, there are significant differences in performance depending on *how* this information is passed to the model: passing the task information directly to the classifier (TIML with no encoder) yields no improvement in F1 score in Brazil relative to the models trained without task information. TIML achieves the best F1 score for this task compared to all other models. The encoder architecture therefore provides a significant boost in performance, yielding the highest mean AUC ROC and F1 scores.

Conclusion

Accurate cropland and crop type maps are important tools in minimizing the worst effects of climate change, but some regions lack the data to develop these maps, particularly when targeting specific crops. Global crop data exists, but learning from this global data to perform well in data-sparse areas can be challenging. We introduce task-informed meta-learning (TIML), a method for conditioning the model with prior information about a specific task. In addition, we introduce the concept of "forgetful" meta-learning, which can improve meta-learning performance when there are many similar tasks to learn from. We find TIML enables the model to learn better features from the global CropHarvest dataset to perform well in a range of agroecologies and dataset-size regimes, outperforming a variety of benchmark models.

All code used to train these models is available at `https://github.com/nasaharvest/timl`.

References

2020. Programme, concepts and definitions. In *World Programme for the Census of Agriculture*, volume 1. FAO.

Collins, L.; Mokhtari, A.; and Shakkottai, S. 2020. Task-Robust Model-Agnostic Meta-Learning. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 18860–18871. Curran Associates, Inc.

Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 1126–1135. PMLR.

Hendrycks, D.; and Gimpel, K. 2016. Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415*.

Jamal, M. A.; and Qi, G.-J. 2019. Task Agnostic Meta-Learning for Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kerner, H.; Tseng, G.; Becker-Reshef, I.; Nakalembe, C.; Barker, B.; Munshell, B.; Paliyam, M.; and Hosseini, M. 2020. Rapid Response Crop Maps in Data Sparse Regions. In *ACM SIGKDD Conference on Data Mining and Knowledge Discovery Workshops*.

Ozdemir, D. 2021. The Impact of Climate Change on Agricultural Productivity in Asian Countries: a heterogeneous panel data approach. *Environ Sci Pollut Res Int*.

Patterson, T.; and Kelso, N. V. ???? Natural Earth. https://www.naturalearthdata.com/.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.

Perez, E.; Strub, F.; De Vries, H.; Dumoulin, V.; and Courville, A. 2018. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Ritchie, H. 2019. Food production is responsible for one-quarter of the world's greenhouse gas emissions. *Our World in Data*.

Rußwurm, M.; Wang, S.; Korner, M.; and Lobell, D. 2020. Meta-learning for few-shot land cover classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 200–201.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1): 1929–1958.

Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; and Zhang, L. 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237: 111322.

Triantafillou, E.; Larochelle, H.; Zemel, R.; and Dumoulin, V. 2021. Learning a Universal Template for Few-shot Dataset Generalization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, 10424–10433.

Tseng, G.; Kerner, H.; Nakalembe, C.; and Becker-Reshef, I. 2021a. Learning To Predict Crop Type From Heterogeneous Sparse Labels Using Meta-Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1111–1120.

Tseng, G.; Zvonkov, I.; Nakalembe, C.; and Kerner, H. 2021b. CropHarvest: a global satellite dataset for crop type classification. In *Neural Information Processing Systems Datasets and Benchmarks Track*.

Vuorio, R.; Sun, S.-H.; Hu, H.; and Lim, J. J. 2019. Multimodal Model-Agnostic Meta-Learning via Task-Aware Modulation. In *Neural Information Processing Systems*.

Wang, A. X.; Tran, C.; Desai, N.; Lobell, D.; and Ermon, S. 2018. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18. ISBN 9781450358163.