HarvestNet 2.0: Compressed CNNs and Augmented Dataset in Harvest Pile Detection

Edrian Liao[1,2*] **Shun Sakai**[1,2*] **Meixiang Du**[3]
[1]Electrical and Computer Engineering [2]Computer Science [3] Rhodes Information Initiative
{edrianpaul.liao, shun.sakai, meixiang.du}@duke.edu

Abstract

Smallholder farming system mapping in rural areas is essential for monitoring food security and agricultural policy. However, current smallholder farming detection models are computationally expensive and data collection for model training is challenging. Building off previous research [1], in this work, we explore the development of a compressed ResNet-50 model for this task and verify and propose a new dataset creation technique (HarvestNet 2.0) for model training. All code used in the experiments is found in this repository [2].

1 Introduction

Smallholder farming systems are the most common form of agriculture in the world, especially in the Global South. [1] These systems are usually owned by a family of farmers, operating under a small-scale agricultural model. Within Sub-Saharan Africa (SSA) and South Asia (SA), 80% of the food consumed comes from these agricultural systems. [3] Despite this, there has been low productivity observed from these farms primarily due to the exacerbating effects of climate change and unfavorable policies. [3] This is precisely why accurately mapping these systems helps monitor food security and improve agricultural extension and development policies. Current approaches to mapping these systems take a lot of time and human labor. Spectral-based classification methods lack precision as the reflectance pattern exhibited by crops is indiscernible from that of weeds and wild vegetation. [1]

Recently, a lab at Stanford proposed the presence of harvest piles as an excellent indicator of smallholder farming activities. [1] Consequently, they focused on utilizing existing neural network architecture, such as ResNet-50, to detect these harvest piles from low-resolution satellite imagery in the Tigray and Amhara regions of Ethiopia. We take this approach further, recognizing (1) the lack of sophisticated computational resources and microsatellites not only in Ethiopia but also in the Global South, and (2) the high costs attributed to labeling the data. [4][5]

Our contributions are as follows:

- We implement pruning, a common model compression technique, that removes unnecessary parameters in a given neural network.
- We attempt to improve the representation power of the model using Dynamic Convolutions.
- We verify the efficacy of ResNet-50 in detecting the small harvest piles within a satellite image.
- We propose a new dataset, HarvestNet 2.0, that reduces the need to procure more ground truth data and access very high-resolution (VHR) satellite imagery.

Final Project Report for ECE 661: Computer Engineering, Machine Learning, and Deep Neural Networks

2 Related Works

HarvestNet [1]. HarvestNet compiled 6915 satellite images from SkySat (database of satellite imagery) from the harvesting season of October to May. These images were normalized, to account for different latitudes and acquisition times and then sharpened and color-corrected to improve visual performance. Then, this data was hand-labeled using a multi-staged committee with experts. After this stage, there were approximately 7k hand-labeled low-resolution images each of size 512x512 pixels. Train and test sets were created from the 7k images with an 80:20 ratio emphasizing the prevention of geographical overlap while maintaining similar geographic distribution *(Appendix A)*.

The group then evaluated preliminary models on their HarvestNet dataset. These models include MOSAIKS, SATMAE, Swin Autoencoder, Saltas, and ResNet-50. Within the models investigated, the Swin Autoencoder and ResNet-50 performed significantly better than their counterparts. The Swin Autoencoder had the highest overall accuracy at 0.8087, followed by ResNet-50 at 0.7918.

Model Compression for Increased Model Accessibility. Rural areas where machine learning and IoT technologies could be deployed lack high-bandwidth connection, bandwidth, power supply and scaling sensors [6]. Thus, an emerging research topic is the development of computationally friendly machine learning models for deployment on IoT devices in rural areas. For cloud computing services, the particular bottlenecks include memory and storage limitations, power consumption, bandwidth, and battery capacity [7]. As such, in an effort to overcome these bottlenecks, the focus for developing such models has been placed on model compression and increasing model computational efficiency.

Linear Interpolation for Upscaling Images. Super-resolution algorithms for small satellite imagery can be divided into three (3) categories: interpolation, pan-sharpening, and CNN. [8]. Within the context of our model simplification approach, interpolation was the best strategy to employ as the rest of the categories incur an algorithm and preprocessing overhead. **Inteprolation** involves assigning values to newly created elements in the matrix when an image is upscaled. The most common interpolation methods are nearest neighbor (pixel duplication), bilinear, and cubic. Each has its own advantages depending on the task at hand. **Bilinear**, specifically, ensures an affine function between pixels as it takes the average of the four neighboring pixels of the original image. This continuity smoothens the edges that can counterintuitively help make the model more robust in detecting harvest piles, as most of the time, the harvest piles appear blended across lands where crops are planted. Eq. (1) shows a function $u(x, y)$ that will transform the original image into an upscaled one.

$$u(x,y) = (1-\langle x\rangle)(1-\langle y\rangle)I^{LR}_{[x],[y]} + x(1-\langle y\rangle)I^{LR}_{[x]+1,[y]} + y(1-\langle x\rangle)I^{LR}_{[x],[y]+1} + \langle x\rangle\langle y\rangle I^{LR}_{[x]+1,[y]+1} \tag{1}$$

, where $I^{LR}_{[x],[y]}$ is the input image matrix.

3 Methodology

3.1 Baseline ResNet-50

To obtain a baseline ResNet-50 model and to verify results from [1], we trained a vanilla ResNet-50 model with weights and parameters given in the paper. The loss function was binary cross entropy and the model was trained using fp16 mixed precision, a one_cycle_lr scheduler with initial learning rate 1e-3, and a MADGRAD optimizer with learning rate 1e-3, momentum 0.9, and weight_decay 0. The model was trained for 30 epochs using the HarvestNet `train.csv` dataset and evaluated on the HarvestNet `test.csv` dataset. Train and test were given in [1] and contained 5531 and 1383 images respectively (one image was broken and thus unused in both our experiments and the original paper).

3.2 Model Compression

To explore model compression, we performed pruning. Pruning by percentage, pruning by percentage with fine-tuning, iterative pruning, and global iterative pruning were investigated using the best baseline ResNet-50 model from training (`best_resnet.pt`). To determine whether a sparse model with similar accuracy performance to the baseline could be found, two investigations were performed. The first evaluates the impact of pruning percentage on accuracy. The second evaluates the impact of pruning method on accuracy. For the former, pruning percentages of 30, 50, 70, and 80 were evaluated using pruning by percentage and global iterative pruning. For the latter, a pruning percentage of 70 was investigated over all four pruning methods.

3.3 Model Improvement

Data Augmentation In addition to model compression, we looked to improve the performance of the overall model. We first looked to explore data augmentation. Keeping all other training parameters the same as the baseline ResNet-50 models, we performed 2 different augmentations. The first was horizontal + vertical flips, and the second was horizontal + vertical flips + random rotation (5 degrees). After training using each of the two augmentation groups, the models were evaluated using the HarvestNet `test.csv` dataset.
Dynamic Convolution Another technique we looked into was dynamic convolution. This technique makes inference costs smaller under the same architecture and boosts performance with marginal increases in computational complexity. Using the method and model structure outlined in [9], we trained the model from 30-100 epochs using 4 different sets of hyperparameters shown in Table 1. Here, we expected to observe a boost in accuracy within an acceptable range of increased parameters through dynamic convolution. We utilized the model structure from Dynamic Convolution Decomposition (DCD)—specifically, Dynamic Convolution Matrix Decomposition. This method, compared to other dynamic approaches like attention-based methods, proves easier to train and requires fewer parameters without sacrificing accuracy.

3.4 HarvestNet 2.0

Dataset Creation. To attain high-resolution imagery from low-resolution imagery, we created a new dataset, HarvestNet 2.0, consisting of 27,660 satellite images scattered across Ethiopia's Tigray and Amhara regions. This is done by subdividing the 7k hand-labeled satellite images in the original HarvestNet dataset into four (4) equally sized patches. These are then subjected to the linear interpolation upsampling algorithm *(Section 2.1)* to bring them back to their original image dimensions. Finally, we fed the patches from the positively labeled images into the baseline ResNet-50 model to generate its new labels. We automatically labeled the patches from the negatively labeled images as negative samples.

ResNet-50's efficacy on harvest pile detection. An incident benefit of creating such a dataset is our ability to prove the efficacy of ResNet-50 in detecting the small harvest piles within the satellite image. A satellite image has a lot of background features that might easily be learned by the neural network. One of the ways to prevent this from happening is to feed the neural network with more negative samples that might discount such features in positively labeled images. HarvestNet 2.0 aids in proving this during the relabeling process of the patches.

Testing improvements on using very high resolution (VHR) satellite imagery. On a more important note, we want to see whether our method of generating high-resolution imagery based on low-resolution imagery improved the performance of the ResNet-50 model. We test our hypothesis by training the ResNet-50 model on HarvestNet 2.0 and testing the model on both HarvestNet and HarvestNet 2.0. Although HarvestNet 2.0 labels were determined by the ResNet-50 model, we argue that doing such an experiment still sheds light on how high-resolution imagery can improve the performance of convolutional neural networks in detecting harvest piles. This is because the model we used in relabeling the patches was trained on HarvestNet while the model we used in testing was trained on HarvestNet 2.0.

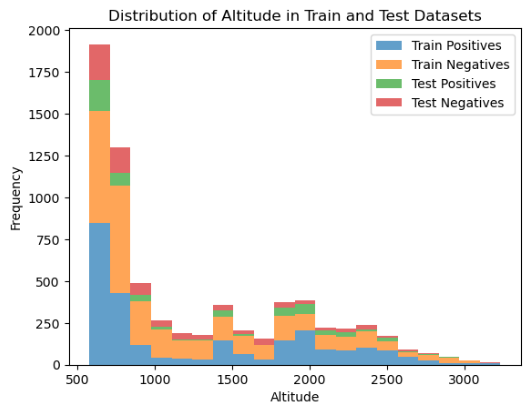
In training HarvestNet 2.0, we used the same parameters. However, instead of using binary cross-entropy loss (BCE loss), we used focal loss that aims to improve the model's performance on hard examples, i.e., positively labeled images. [10] This was due to the severe imbalance within HarvestNet 2.0 that occurred when the positively labeled images were further divided into negatively and positively labeled patches. Moreover, we upsampled the minority samples (positive) by weighing them heavier during the data sampling process. [11]

4 Experiments

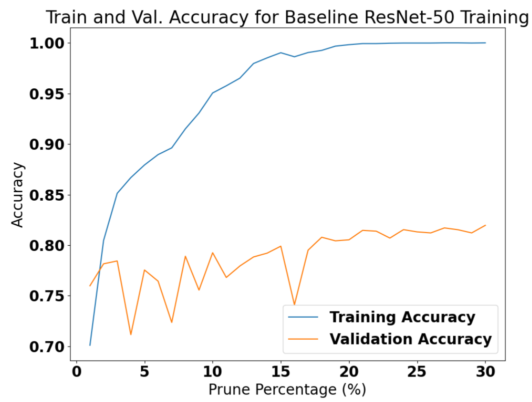
4.1 Baseline ResNet-50

This experiment looks to create a baseline ResNet-50 model that is similar to the benchmark model in the original paper [1]. A ResNet-50 model was trained three times using the outlined methodology and the best performance was taken for the baseline. The best-performing model (highest overall

accuracy) was saved as `best_resnet.pt` and used in the following investigations. The highest-performing model achieved 0.8110 accuracy on the HarvestNet `test.csv` dataset which was slightly higher than the performance observed in the Stanford Paper (0.7918).
Another observation we noted was the convergence and overfitting of the model. In our investigation, the ResNet-50 model started to converge around epoch 18 whereas the Stanford paper observed convergence around epoch 15. Furthermore, we noted the ResNet-50 model to be overfitting, with around a 18% performance difference between train and test accuracy while training, shown in Fig 1 (b). We conclude that this is mainly the result of the class imbalance in the HarvestNet dataset, shown in Fig (a), and ResNet-50 being computationally complex for this task.



(a) Distribution of Data in HarvestNet Dataset



(b) Performance of the Baseline ResNet-50 model

Figure 1: Data Imbalance and Baseline Metrics

4.2 Model Compression

This experiment looks to determine whether the baseline ResNet-50 model can be compressed without losing accuracy performance. To start, different pruning percentages were evaluated to determine an optimal pruning percentage. The results for both pruning by percentage and global iterative pruning are shown in Appendix B: Tables 3 and 4. For pruning by percentage, as the percentage pruned increases, there is a significant decrease in accuracy (from 0.7965 at 30% to 0.4413 at 80%). However, using global iterative pruning, the accuracy is maintained near the baseline model accuracy from 30% to 70% pruning. From this, we concluded that using basic percentage-based pruning, accuracy decreases with increased prune percentage. However, we can achieve similar accuracy to the baseline model with up to 70% sparsity using global iterative pruning.

To further investigate model compression, we evaluated the four different pruning methods at 70% sparsity (the highest sparsity which maintained baseline accuracy in the previous investigation). The results of this investigation can be seen in Appendix B: Table 5. From these results we can observe that the model at 70% sparsity generally maintains its accuracy for pruning methods that contain some aspect of retraining (FineTine, Iterative Pruning, and Global Iterative Pruning). Further, the best-performing model had an accuracy of 0.8068, which is only 0.0042 off from the baseline model performance. Thus, we can conclude that we can significantly compress the baseline model (up to 70% sparsity) and still achieve results equivalent to that of the baseline ResNet-50 model.

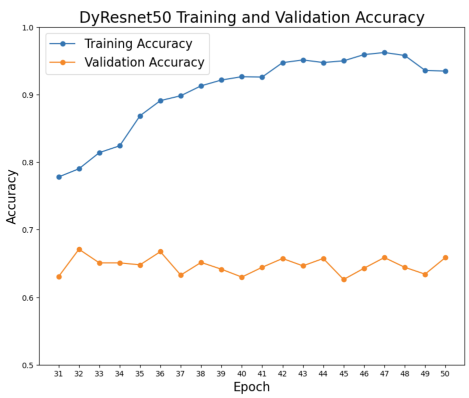
4.3 Model Improvement

Data Augmentation. For ResNet-50 with horizontal and vertical flips, we observed 0.8232 accuracy on the test set, a 1.2% improvement from the baseline ResNet-50. For the ResNet-50 with horizontal and vertical flips + random rotation, we observed a similar performance increase at 0.8217 accuracy (Appendix D: Table 6, Figure 9). To verify similar performance increases on a compressed model, we also evaluated data augmentations with 70% global-iterative pruning. This pruned ResNet-50 model achieved 0.8182 and 0.8239 accuracy on the HarvestNet test set (for augmentation set 1 and 2 respectively). The best compressed model is a 1.3% increase from the baseline ResNet-50 model, confirming the development of a compressed ResNet-50 with increased performance from the baseline model.

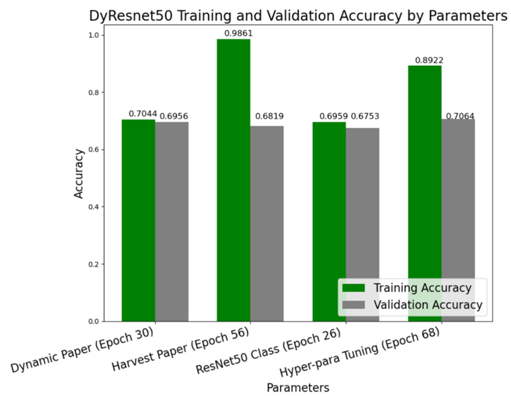
Dynamic Convolution. Despite training the model ranging from 30-100 epochs using four different sets of hyperparameters, including those from the original paper, we only achieved a validation accuracy of 70.6%. Even after 30 epochs, the model with all four sets of parameters tends to overfit, with training accuracy reaching as high as 99%, while validation accuracy plateaued at 70%.

We employed the best-performing model weights from the previous training round as our starting point. Then, we conducted aggressive hyperparameter tuning using four different parameter settings, including parameters inherited from [1] as well as the original dynamic conclusion via [9].

However, after epoch 30, while the training accuracy continued to grow, the validation accuracy started to plateau at 70%, demonstrating overfitting. After aggressive hyperparameter tuning, including training the model for up to 100 epochs at times, the improvement was marginal. The best validation score achieved was 0.7064 after increasing the L2 regularization from 1e-4 to 1e-3, while the training accuracy reached 0.8922.



(a) Convergence of DyResNet-50 from 31-50 epochs



(b) Performance from different hyperparameters

Figure 2: DyResNet-50 Overall Performance

Table 1: Different sets of hyperparameters used in DyResNet-50

Hyperparameter	[9]	[1]	ResNet-50 (from class)	Tuned Hyperparameters
Initial LR	0.1	1e-3	0.01	1e-2
Momentum	0.9	0.9	0.9	0.9
L2-Regularizer	1e-4	1e-4	1e-4	1e-3
Optimizer	SGD	Adam	SGD	SGD
Scheduler	StepLR	OneCycleLR	-	-

Observations suggest that the model may be overly complex for the task at hand. Vanilla ResNet-50 has already demonstrated signs of overfitting, and the dynamic model only worsened the situation. It exhibited a greater capacity to memorize the training data but struggled to generalize to unseen cases. Additionally, the limited variety like the task data may have hindered the full utilization of the adaptive inference power.

4.4 HarvestNet 2.0

ResNet-50's efficacy on harvest pile detection. Four (4) satellite images were closely studied. Once partitioned into patches, we tested them on the baseline ResNet-50 model. Appendix C: Figure 8 shows the predictions made by the model. Only the patches with the harvest piles had a softmax score of ≥ 0.5 which means that these figures are labeled positive. Appendix C (a) shows that only the top left patch contains the harvest piles while (b) shows that only the bottom two patches contain the harvest piles. These can be confirmed just by looking at the photos. On the other hand, all patches from the negatively labeled images had a softmax score of approximately 0 which means that these figures are labeled negative. These results show the efficacy of the ResNet-50 model in learning the presence of harvest piles within the satellite image and not the background features that could easily distort the CNN's predictions.

Testing improvements on using very high resolution (VHR) satellite imagery. After the generation of patches from the satellite image, we created new CSV files to account for these new images from HarvestNet. Once both images and the metadata were ready, we trained a ResNet-50 model on HarvestNet 2.0, i.e., the upscaled patches for 30 epochs. Such a model was then evaluated on both HarvestNet and HarvestNet 2.0. Table 2 shows the performance of the model. The model borne from HarvestNet 2.0 performed better on the images in HarvestNet 2.0 (approx. 93%) than on the images in HarvestNet (approx. 77%). This is due to the fact that the harvest piles found in HarvestNet 2.0 are visually bigger than those found in HarvestNet.

This presents a weak proof that working with VHR satellite imagery may improve the performance of convolutional neural networks like ResNet-50. These results also show that our method of subdividing and linear interpolation upscaling can reduce the need for state-of-the-art expensive satellites that capture high-resolution satellite imagery. Moreover, we can see that we successfully minimized the drop in precision and recall scores. This suggests that upsampling the minority group and using focal loss are both effective strategies in dealing with unbalanced datasets.

Table 2: Performance of ResNet-50 on HarvestNet and HarvestNet 2.0 **upsampled and using focal loss, **shuffled and using cross-entropy loss*

Experiment	Accuracy	Precision	Recall
Baseline (Trained and tested on HarvestNet)	0.8110	0.8394	0.7069
Trained on HarvestNet 2.0 and tested on HarvestNet*	0.7748	0.8763	0.5650
Trained and tested on HarvestNet 2.0**	0.9327	0.7117	0.6229

5 Discussion and Conclusion

In summary, we made contributions to the training and deployment of machine learning models for use in small farmholder systems detection. A compressed baseline ResNet-50 model was developed achieving 0.8068 accuracy at a 70% reduction of weights. Further, when paired with data augmentation, we achieved 0.8239 accuracy at a 70% reduction of weights: a 1.3% accuracy increase over the baseline ResNet-50 model. Not only does this achieve successful compression of the model but also diminishes its computational complexity, facilitating broader application in the developing world. In terms of data collection, a method to increase harvest data using image division and upscaling was proposed. Using this improved dataset (HarvestNet2.0), a baseline ResNet-50 model achieved 93% accuracy on the HarvestNet 2.0 test set and 77% on the HarvestNet test set. This presents a weak proof that using VHR satellite imagery improves CNN performance such as ResNet-50. Further proving this involves hand-labeling the patches.

In the future, we look to experiment with neural network architectures such as ResNet-20 to determine whether less complex models may be more suitable for this task. Lighter models such as MobileNet or EfficientNet could be considered, especially for deployment in resource-constrained environments. We also consider employing more advanced data augmentation. While basic transformations like flipping and rotation have proven effective, incorporating more sophisticated augmentation techniques such as scaling, cropping, and color adjustments may provide further gains in model robustness. We also plan to explore utilizing transfer learning, leveraging pre-trained models on similar tasks, and fine-tuning them to the specific dataset which could reduce training time and computational resources while maintaining or improving model accuracy.

References

[1] Jonathan Xu et al. *HarvestNet: A Dataset for Detecting Smallholder Farming Activity Using Harvest Piles and Remote Sensing*. Aug. 23, 2023. DOI: 10.48550/arXiv.2308.12061. arXiv: 2308.12061[cs]. URL: http://arxiv.org/abs/2308.12061 (visited on 03/04/2024).

[2] Edrian Liao. *edrian-liao/harvestnet*. original-date: 2024-04-08T12:29:33Z. Apr. 29, 2024. URL: https://github.com/edrian-liao/harvestnet (visited on 04/29/2024).

[3] Nurudeen Abdul Rahman et al. "Editorial: Sustainable intensification of smallholder farming systems in Sub-Saharan Africa and South Asia". In: *Frontiers in Sustainable Food Systems* 8 (Apr. 9, 2024). Publisher: Frontiers. ISSN: 2571-581X. DOI: 10.3389/fsufs.2024.1399430. URL: https://www.frontiersin.org/articles/10.3389/fsufs.2024.1399430 (visited on 04/28/2024).

[4] *AI in the Global South: Opportunities and challenges towards more inclusive governance*. Brookings. URL: https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance/ (visited on 04/28/2024).

[5] *Contours of Space Diplomacy in the Global South | Science & Diplomacy*. May 16, 2023. URL: https://www.sciencediplomacy.org/perspective/2023/contours-space-diplomacy-in-global-south (visited on 04/28/2024).

[6] Miguel A. Guillén et al. "Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning". In: *The Journal of Supercomputing* 77.1 (Jan. 1, 2021), pp. 818–840. ISSN: 1573-0484. DOI: 10.1007/s11227-020-03288-w. URL: https://doi.org/10.1007/s11227-020-03288-w (visited on 04/29/2024).

[7] Phumlani T. Simelane, Okuthe P. Kogeda, and Manoj Lall. "A cloud computing augmenting agricultural activities in Marginalized Rural Areas: A preliminary study". In: *2015 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*. 2015 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC). May 2015, pp. 119–124. DOI: 10.1109/ETNCC.2015.7184820. URL: https://ieeexplore.ieee.org/abstract/document/7184820 (visited on 04/29/2024).

[8] Kinga Karwowska and Damian Wierzbicki. "Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), pp. 3292–3312. DOI: 10.1109/JSTARS.2022.3167646.

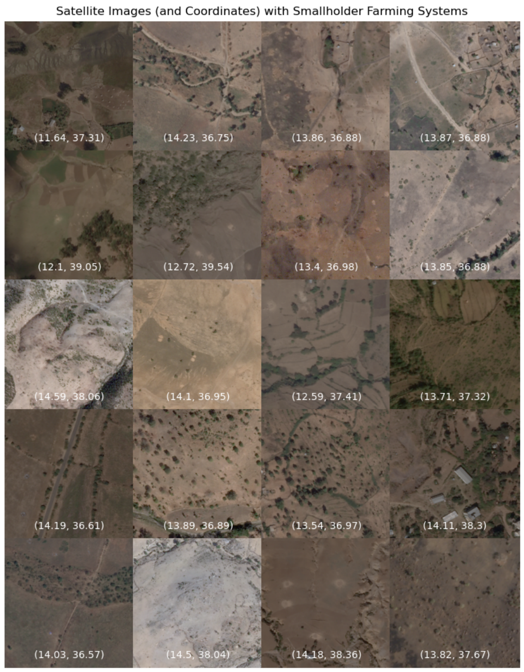
[9] Yunsheng Li et al. *Revisiting Dynamic Convolution via Matrix Decomposition*. arXiv.org. Mar. 15, 2021. URL: https://arxiv.org/abs/2103.08756v1 (visited on 04/29/2024).

[10] Tsung-Yi Lin et al. *Focal Loss for Dense Object Detection*. Feb. 7, 2018. DOI: 10.48550/arXiv.1708.02002. arXiv: 1708.02002[cs]. URL: http://arxiv.org/abs/1708.02002 (visited on 04/29/2024).

[11] Ming. *ufoym/imbalanced-dataset-sampler*. original-date: 2018-05-29T02:15:17Z. Apr. 28, 2024. URL: https://github.com/ufoym/imbalanced-dataset-sampler (visited on 04/28/2024).

A HarvestNet and HarvestNet 2.0 Data Exploration

Below are some examples of positively and negatively labeled satellite images from both datasets.



(a) Positive samples from HarvestNet

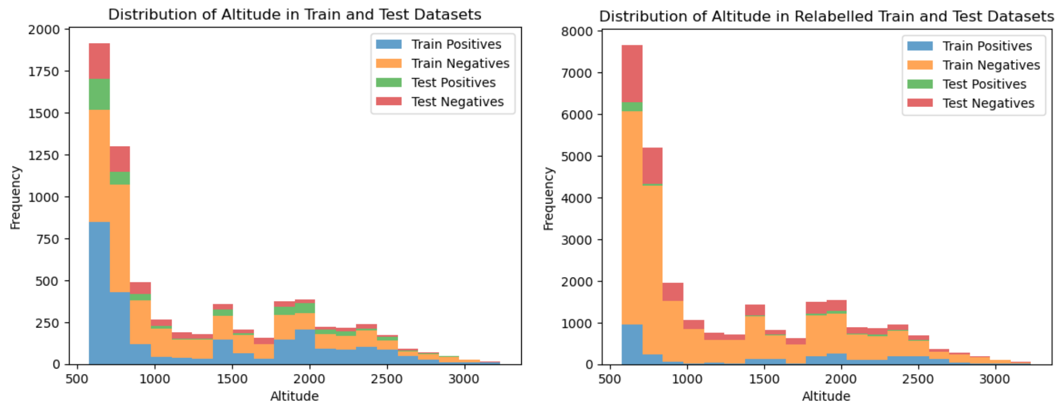
(b) Negative samples from HarvestNet

(c) Positive samples from HarvestNet 2.0

(d) Negative samples from HarvestNet 2.0

Figure 3: Positive and negative samples from HarvestNet and HarvestNet 2.0

Fig. 4 shows the greater imbalance between positive and negative samples in HarvestNet 2.0



(a) HarvestNet

(b) HarvestNet 2.0

Figure 4: Distribution of positive and negative samples on both datasets

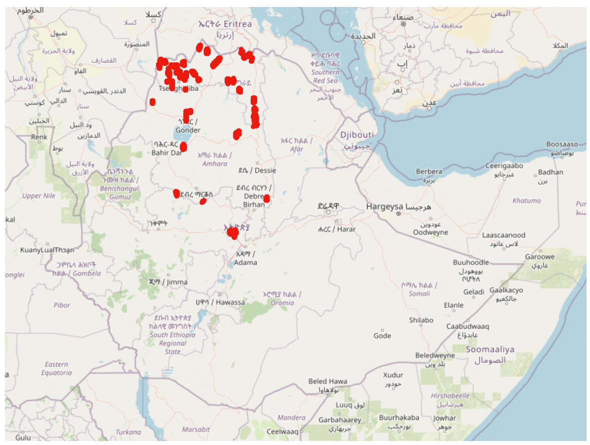


Figure 5: Coordinates within HarvestNet pinned on the map of Ethiopia

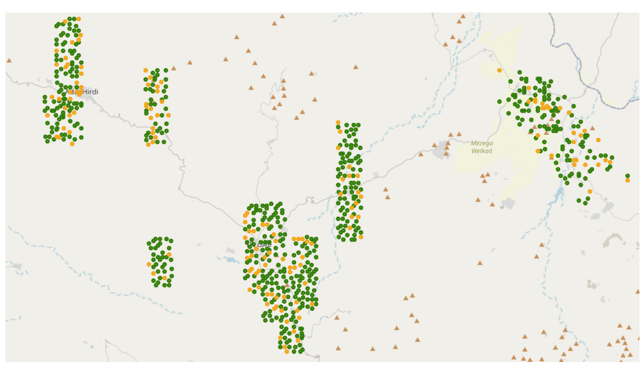
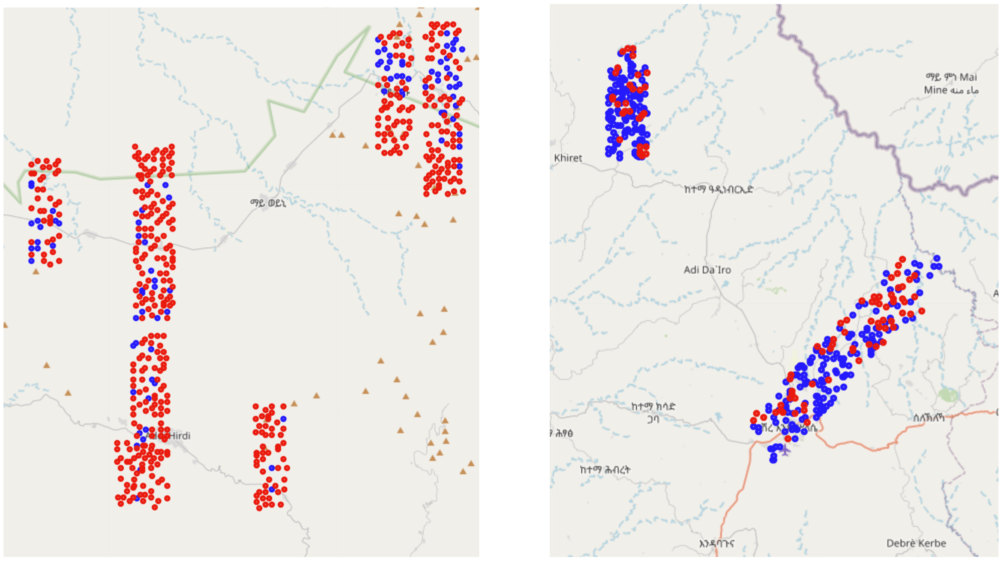


Figure 6: Consistent train (green) and test (yellow) data splits within clusters/regions in Ethiopia



(a) Regions with more positive samples

(b) Regions with more negative samples

Figure 7: Distribution of positive and negative samples with clusters/regions

B Summary of Results during the Model Pruning Experiments

Table 3: Accuracy vs. Pruning Percentage for Pruning by Percentage

Pruning Percentage	Accuracy	F1	AUROC	Precision	Recall
Baseline	0.8110	0.7574	0.8806	0.8394	0.7069
30%	0.7965	0.7487	0.8817	0.6749	0.8076
50%	0.7495	0.8368	0.6749	0.8076	0.8847
70%	0.4441	0.5772	0.4302	0.4315	0.8847
80%	0.4413	0.6029	0.5993	0.4373	0.9900

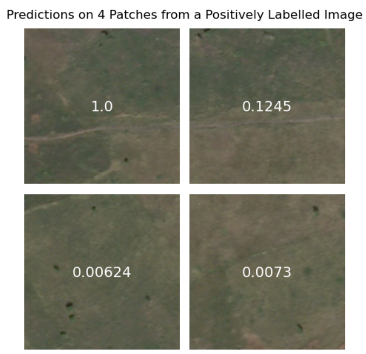
Table 4: Accuracy vs. Pruning Percentage for Global Iterative Pruning

Pruning Percentage	Accuracy	F1	AUROC	Precision	Recall
Baseline	0.8110	0.7574	0.8806	0.8394	0.7069
30%	0.8127	0.7550	0.8783	0.8225	0.7101
50%	0.8064	0.7526	0.8811	0.8445	0.6950
70%	**0.8068**	0.7466	0.8874	0.8443	0.6792
80%	0.4402	0.6053	0.5000	0.4402	1

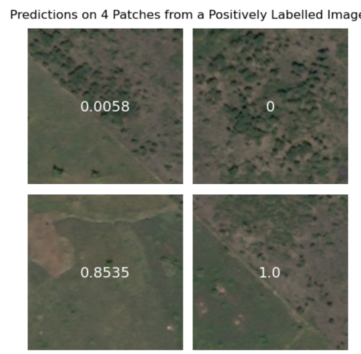
Table 5: Accuracy vs. Pruning Method at 70% Pruning Percentage

Pruning Method	Accuracy	F1	AUROC	Precision	Recall
Baseline	0.8110	0.7574	0.8806	0.8394	0.7069
Pruning by Percentage	0.4441	0.5772	0.4302	0.4315	**0.8847**
Pruning by Percentage + FineTune	0.8050	**0.7522**	0.8808	0.8338	0.6953
Iterative Pruning	0.7978	0.7359	0.8750	0.8265	0.6737
Global Iterative Pruning	**0.8068**	0.7466	**0.8874**	**0.8443**	0.6792

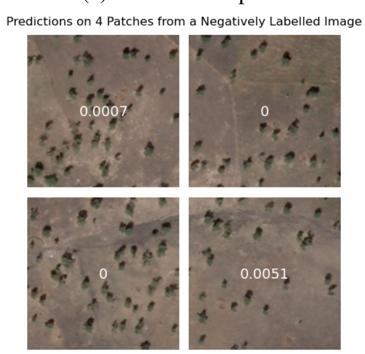
C Relabeling Results for the Patches



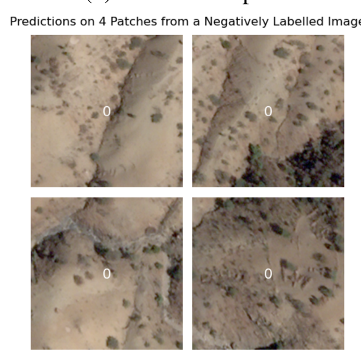
(a) Positive Sample 1



(b) Positive Sample 2



(c) Negative Sample 1



(d) Negative Sample 2

Figure 8: Predictions of ResNet-50 model on patches from both positive and negative samples

D Data Augmentation Results

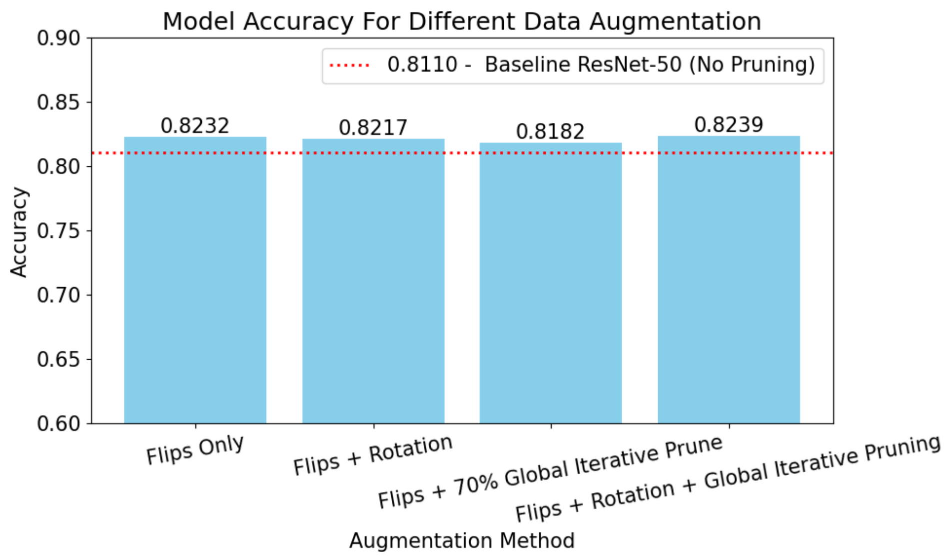


Figure 9: Accuracies for Different Data Augmentation Sets

Table 6: Accuracy vs. Data Augmentation

Method	Accuracy	F1	AUROC	Precision	Recall
Baseline	0.8110	0.7574	0.8806	0.8394	0.7069
Flips Only (Set 1)	0.8232	0.7885	0.8986	0.8092	0.7782
Flips + Rotation (Set 2)	0.8217	0.7885	0.8986	0.8092	0.7782
Set 1 + 70% Global Iterative Pruning	0.8182	0.7707	0.8993	0.8461	0.7164
Set 2 + Global Iterative Pruning	**0.8239**	0.7844	0.9008	0.8070	0.7721