Semi-Supervised Active Learning Framework for Unbalanced, Noisy Datasets

Matthew Bundas
Master's Research Project Final Exam
Adv: Dr. Parth Nagarkar

Introduction To GLOBE

- A NASA-sponsored citizen-science project as means to collect data via an app

- Users take pictures of their surroundings, have option to provide information about observations

- Allows for mass-collection of data throughout the world

- Encourages citizens to engage in science.

https://earthobservatory.nasa.gov/blogs/earthmatters/tag/land-cover/

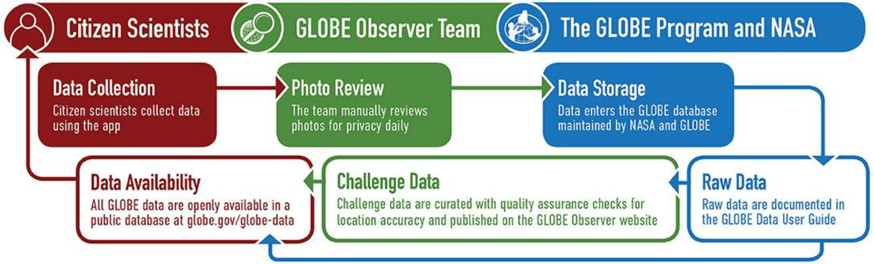
https://observer.globe.gov/about/get-the-app

Introduction To GLOBE - Contribution To Science

Land cover impacts:
- Animal Habitats
- Flooding
- Water Quality
- Landslides
- Forest Fires
- Earth-cycles
- Temperature

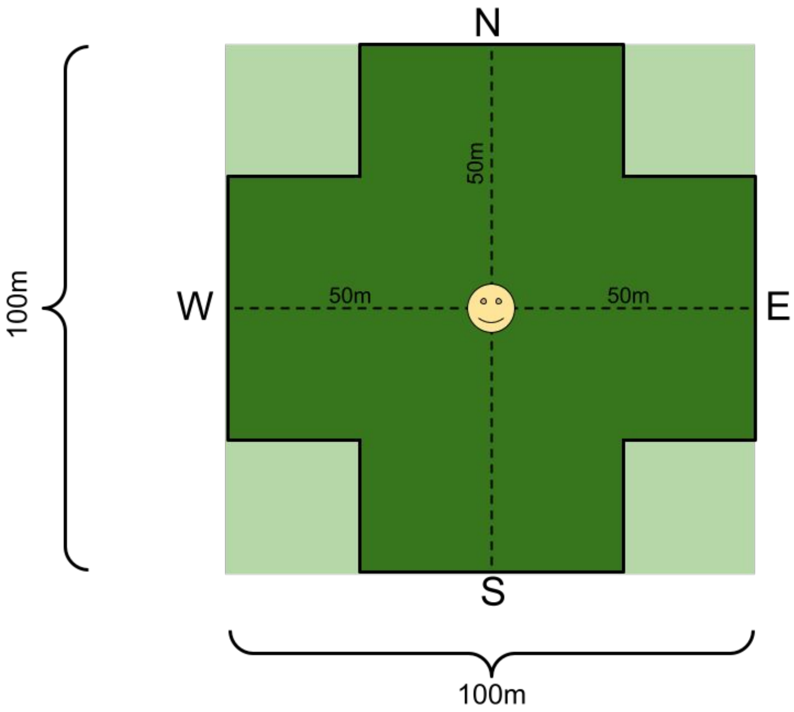
GLOBE data's major contribution is confirming and filling in the gaps of satellite imagery.

https://www.youtube.com/watch?v=vHmwHBFHCQE&t=204s



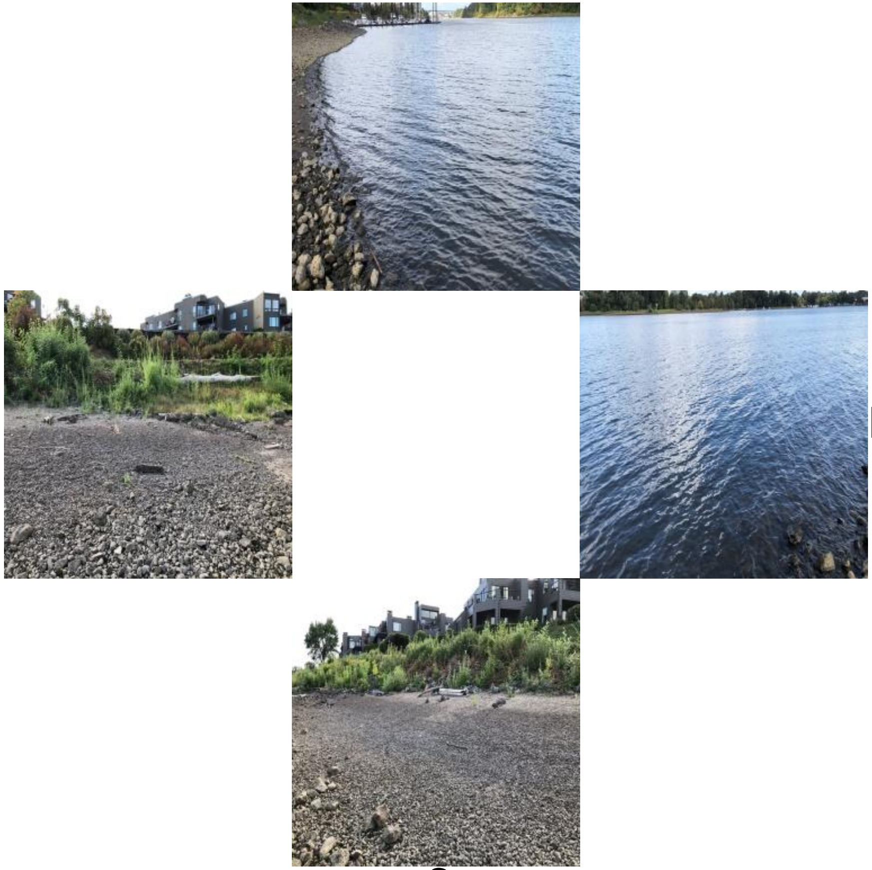
https://www.frontiersin.org/articles/10.3389/fclim.2021.620497/full

Land Cover Observations - Images



N

W



E

S

[1]

Modified UNESCO Classification (MUC) Code Labels

Leading Digit	Land Cover
0	Closed Forest
1	Woodland
2	Shrubland/Thicket
3	Dwarf Shrubland/Thicket
4	Herbaceous Veg.
5	Barren
6	Wetland
7	Open Water
8	Cultivated Land
9	Urban

First Two Digits	Land Cover
01	Evergreen
02	Deciduous
03	Xeromorphic

First Three Digits	Land Cover
021	Tropical
022	Cold w/ Evergreen
023	Cold w/o Evergreen

- Users provide percentage-wise composition of images

- Percentage-wise compositions are averaged among the images, single MUC code is derived for the observation

Examples of Classes

0 - Closed Forest

1 - Woodland

2 - Shrubland/Thicket

3 - Dwarf Shrubland/Thicket

4 - Herbaceous Vegetation

5 - Barren

6 - Wetland

7 - Open Water

[1]

8 - Cultivated Land

9 - Urban

Overview Of Project

Goal: Given that classifying data is optional for users, produce system that can provide classifications for observations

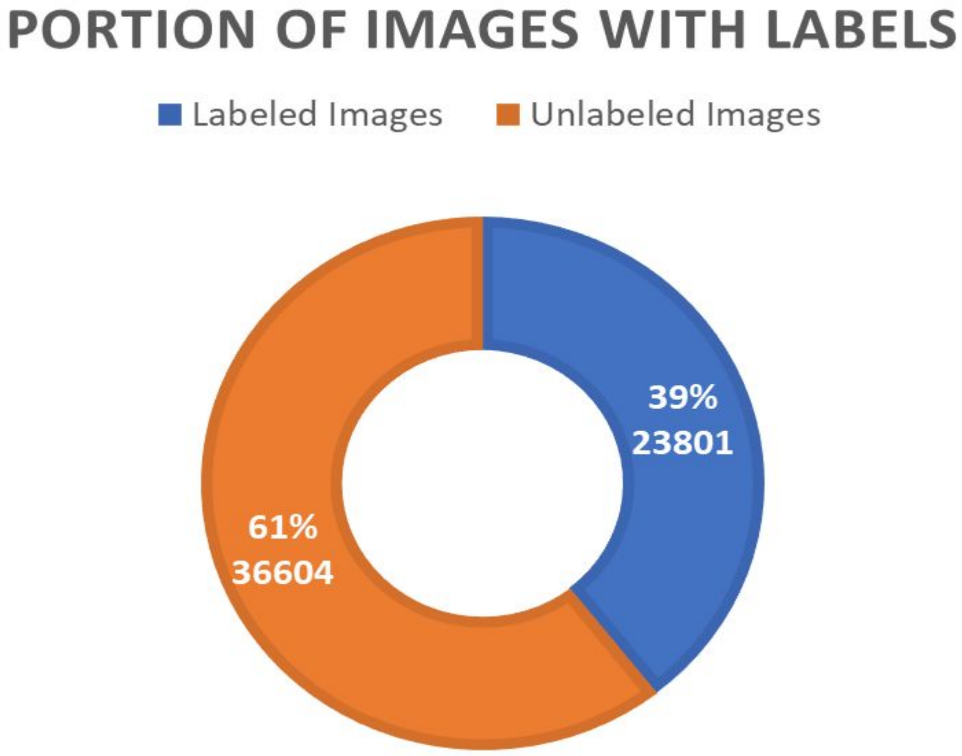
https://www.pinclipart.com/maxpin/woxwT/

Tools Provided: 1) GLOBE Dataset
2) "Hackathon" to help label images

Constraints On Problem: 1) Consider only single images
2) Consider only the primary (first digit) MUC code

General Solution: Use Machine Learning, data science techniques to construct model which can label unclassified images despite challenges.

Dataset and Its Issues - Lack of Labels



Total Number of Images - 60,405

Total Number of Labeled Images - 23,801

61% of images are not labeled

In general it's believed dataset size is positively correlated with performance [5-8]

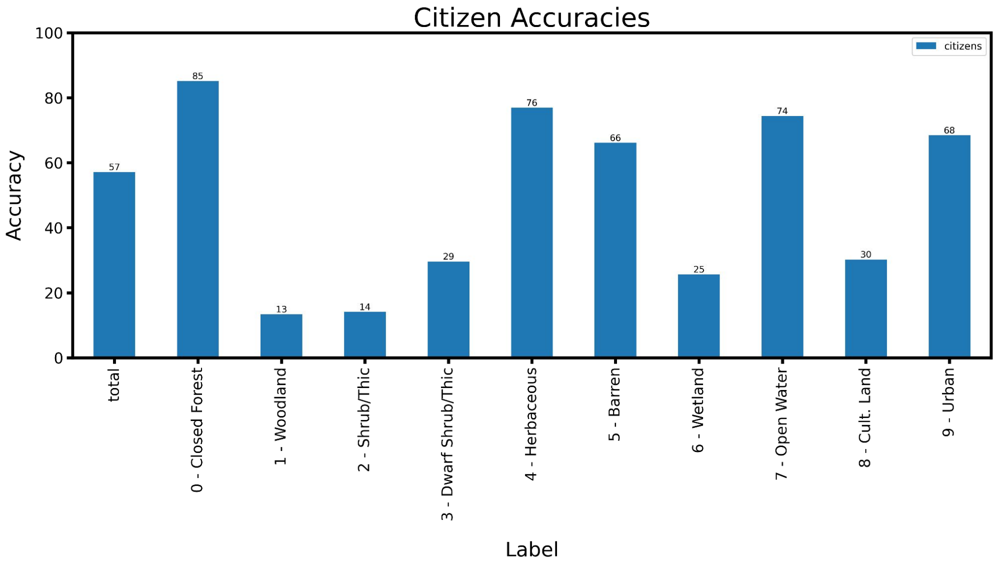
Dataset and Its Issues - Poor Images

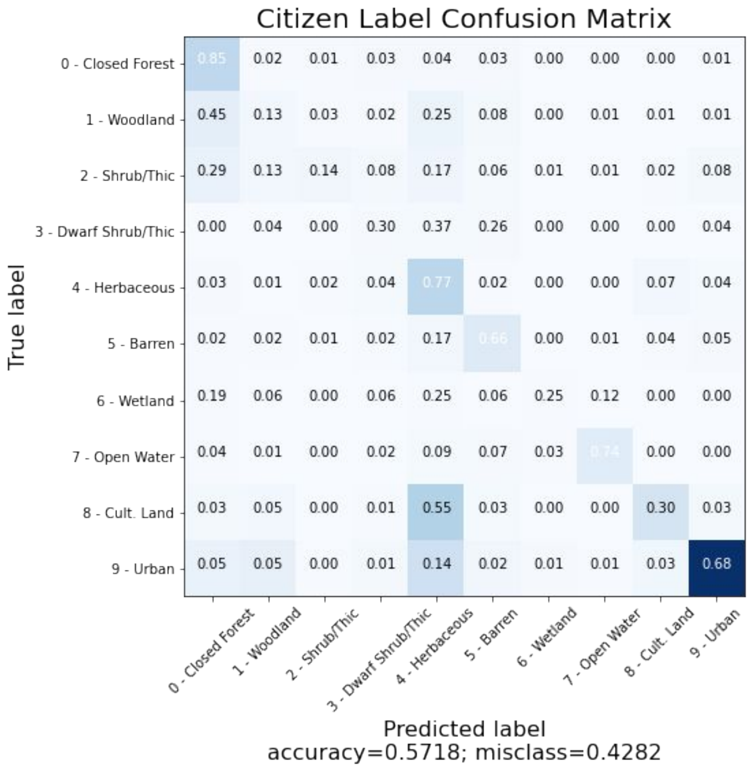
[1]

~2% of images don't have trustable land cover content.

Many more are less than ideal images.

Dataset and Its Issues - Inaccurate Labels

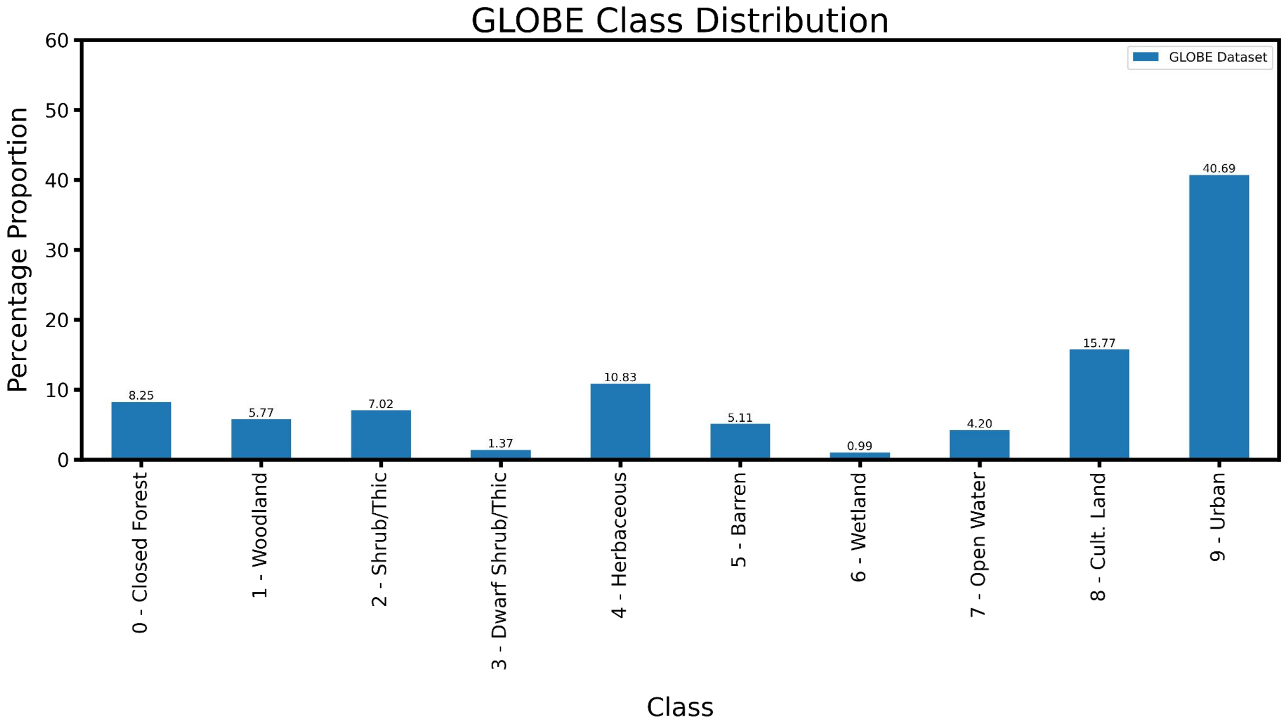




Citizen Accuracy - 57%

Noisy labels can have disastrous effect on model performance [9-10]

Dataset and Its Issues - Class Imbalance



Most Common Class to Least Common Class Ratio - 41

Imbalance in training data causes reduced performance, bias towards the majority classes [11-14]

Problem Statement

Challenges:

1) Missing Labels
2) Label Noise
3) Class Imbalance
4) Image Noise
5) We have more than one challenge

Given the challenges the GLOBE dataset poses, how can we overcome them and create a system that produces meaningful predictions? How can we intelligently use the resources we do have?

Related Work

1) Land Cover and Image Classification in General.

Xu et. al IEEE 2020 [3], Lamas et. al Neuro. 2021 [15], Maggiori et. al IEEE 2017 [16], Buscombe & Ritchie Geo. 2018 [17], Ge et. al GEC 2020 [18]

2) Low Sample Size

Perez et. al 2017 [19], Goodfellow 2016 [20], Liu & Deng [21], Zhao AIp 2017 [22], Hussain et. al UKCI 2018 [23], Ranganthan et. al ICIP 2017 [24]

3) Label Noise

Tanaka et al. IEEE 2018 [25], Veit et al IEE 2017 [26], Han et al IEEE 2019 [27] Lee, Dong-Hyun ICML 2013 [28], Lukasik PMLR 2020 [29]

4) Class Imbalance

Dong et al IEEE 2019 [30], Ando et al 2017 [31], Lin et al AAAI 2018 [32], Johnson et al 2019 [33], Lin et al IEEE 2017 [34]

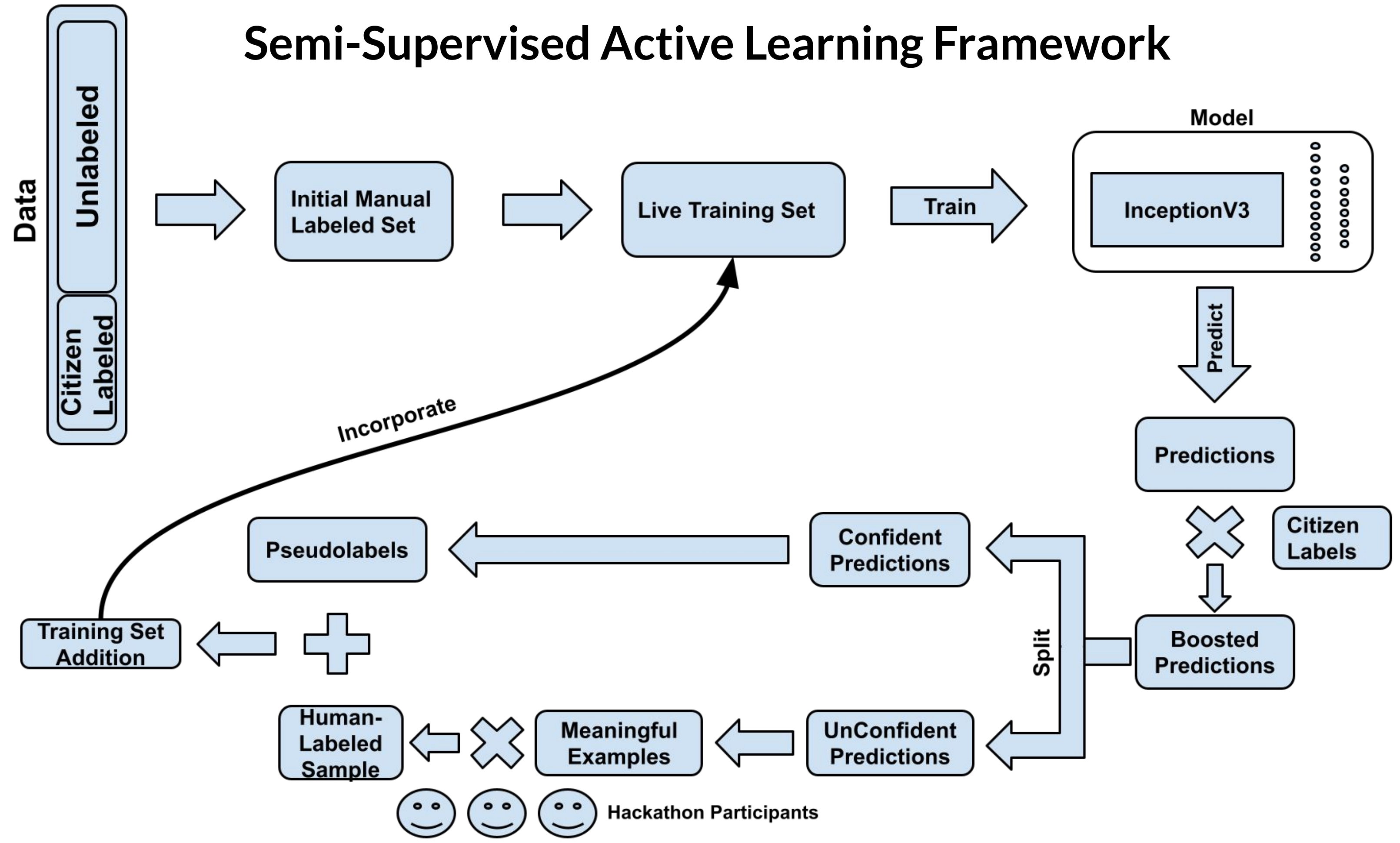
5) Several issues together

Ren et al PMLR 2018 [35], Koziarski et al KBS 2020 [36], Zhong et al IEEE 2019 [37], Chen et al IS 2021 [38], Chen et al Digit. Med. 2019 [39]

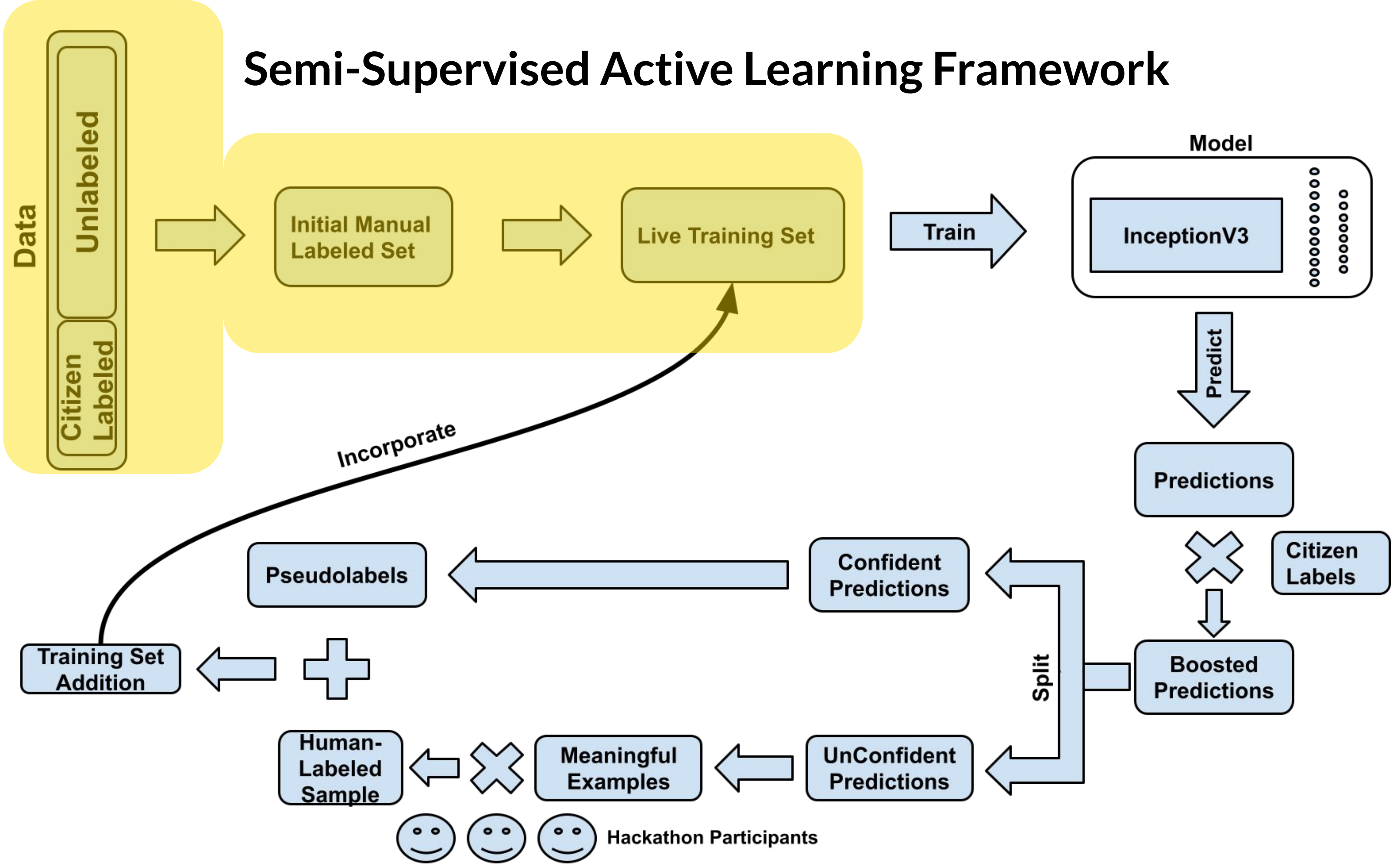
Our Solution In Words

- Use CNN as classifier
- Manually label a small subset of data to serve as clean base
- Grow training dataset iteratively with:
    - Semi-Supervised learning to generally increase training set size with aid of noisy citizen labels
    - Active Learning to select the most useful unlabeled images, be efficient with manual labor
- Apply transfer learning to aid with small sample set
- Use general best Deep Learning practices like Dropout, Image Augmentation, Label Smoothing etc.
- Add 11th class “Class 10” for very bad images

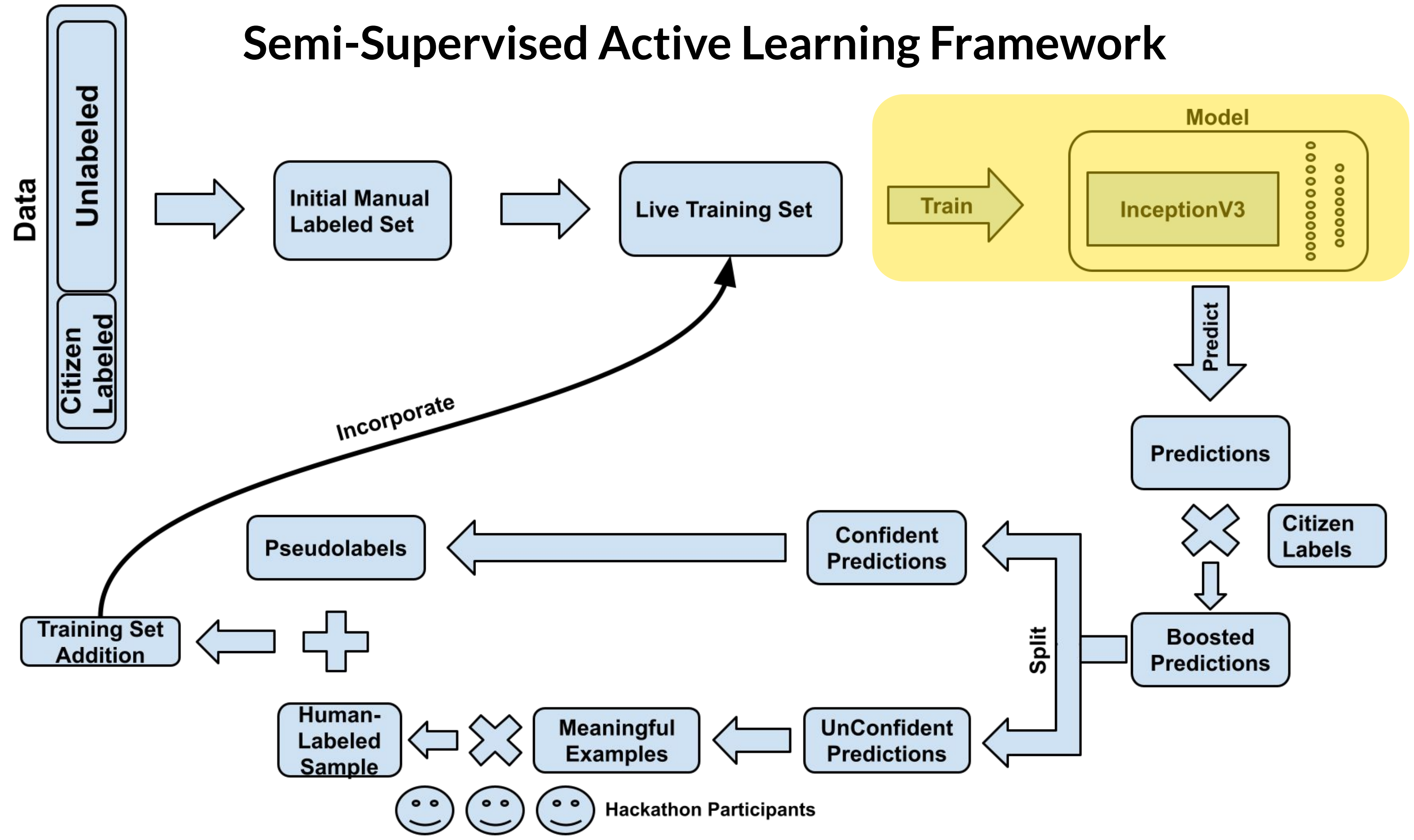
Semi-Supervised Active Learning Framework



Semi-Supervised Active Learning Framework



Semi-Supervised Active Learning Framework

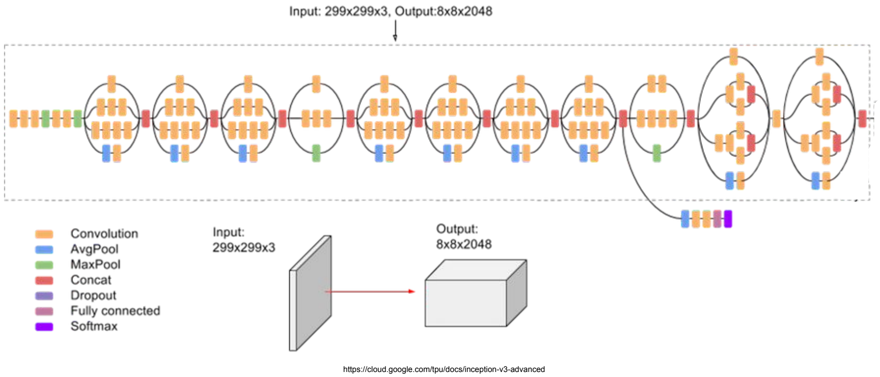


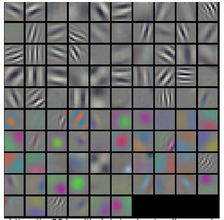
Train the model - Transfer Learning

[1]

Input

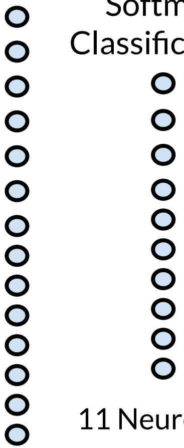
Inception-V3





Fully - Connected Layer

Softmax Classification



0.01	0
0.01	1
0.00	2
0.01	3
0.02	4
0.03	5
0.01	6
0.00	7
0.00	8
0.93	9
0.00	10

11 Neurons

516 Neurons

Feature Extraction

Classification

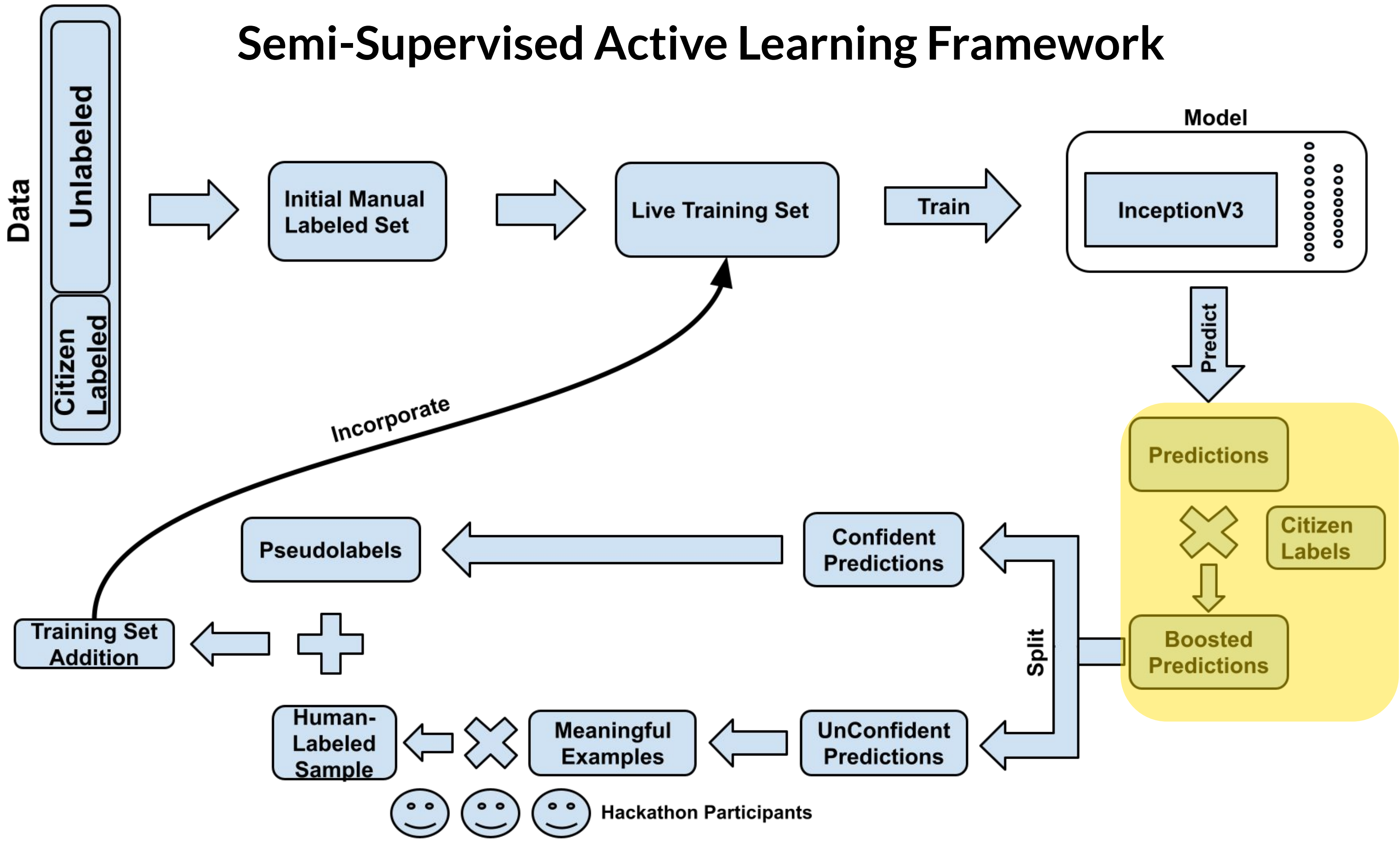
Transfer learning allows us to leverage an external model which has already been trained on millions of images without having to train it ourselves.

All we need to do is fine tune classification layers for our application.

Inception-V3 provides a good balance regarding model size, accuracy, and inference time.

It has been used in similar domains [3,40]

Semi-Supervised Active Learning Framework



Make Predictions - Boosting Prediction Confidence

At every iteration, we assess the accuracy of the citizen predictions, and use their accuracy as a supplemental prediction to enhance the models predictions with formula:

$$P_j[i] = cit\_acc[i] * cit\_weight + P[i] - (cit\_acc[i] * cit\_weight * P[i])$$

i is the class predicted by citizens

P_j is the list containing probabilities image j belongs to each class

$P_j[i]$ is the predicted model probability of class i

$cit\_acc[i]$ is the citizen accuracy of class i

cit_weight is a hyperparameter (0,1) for the amount of weight to put into citizen predictions

Make Predictions - Boosting Prediction Confidence

$$P_j[i] = cit\_acc[i] * cit\_weight + P[i] - (cit\_acc[i] * cit\_weight * P[i])$$

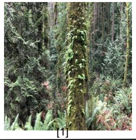
cit_acc - [0.85, 0.13, 0.14, 0.29, 0.76, 0.66, 0.25, 0.74, 0.30, 0.68, 0]

cit_weight = 0.75

[1]

Citizen Prediction - 0
Model Probabilities - [**0.88,** 0.01, 0.02 ... 0.01, 0]
Model Prediction - 0 with probability 0.88

Enhanced Model Probabilities - [**0.96,** 0.01, 0.02 ... 0.01, 0]
Enhanced Model Prediction - 0 with probability 0.96



[1]

Citizen Prediction - 9
Model Probabilities - [0.88, 0.01, 0.02 ... **0.01**, 0]
Model Prediction - 0 with probability 0.88

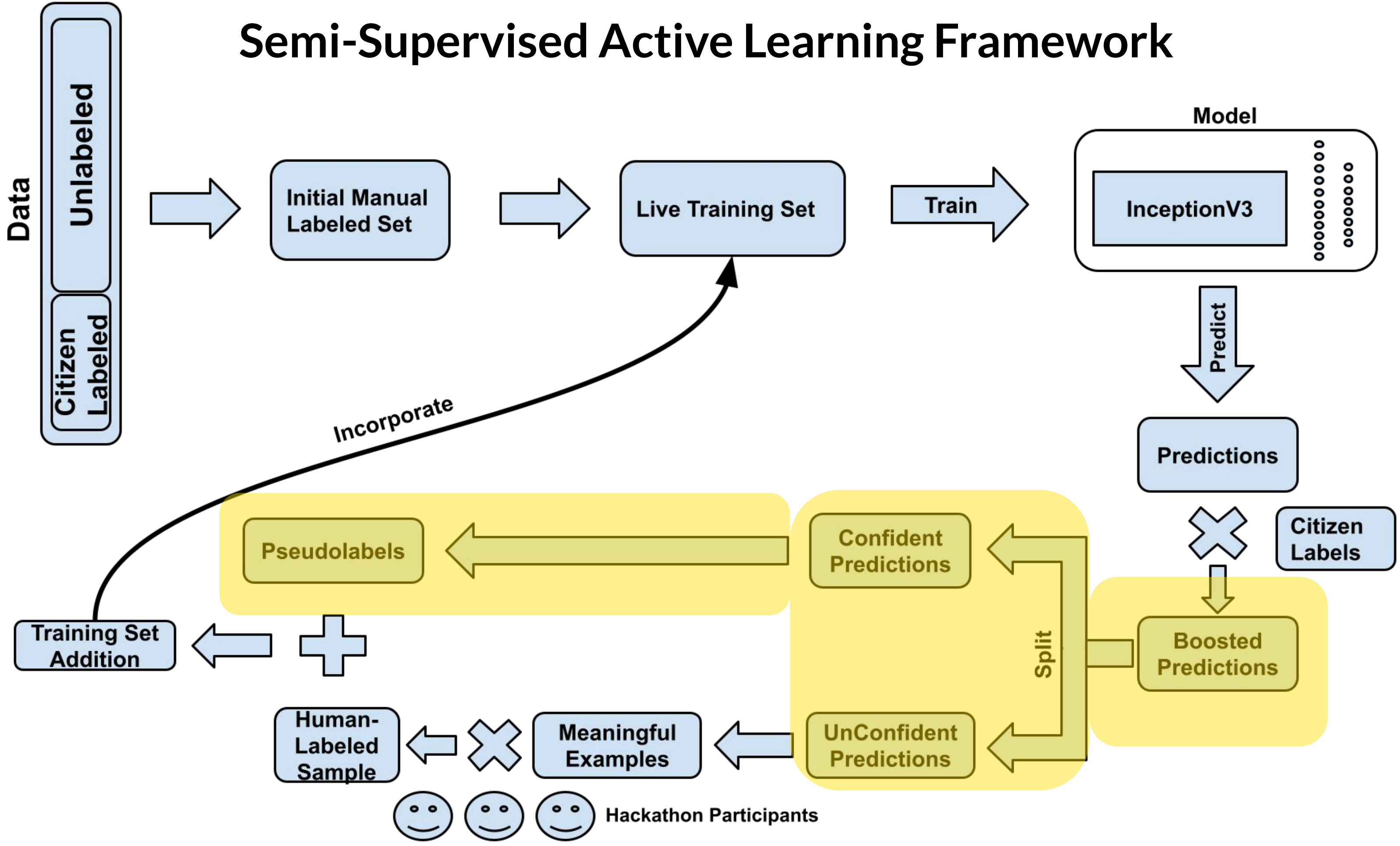
Enhanced Model Probabilities- [0.88, 0.10, 0.02 ... **0.51**, 0]
Enhanced Model Prediction - 0 with probability 0.88

Model Probability	Citizen Accuracy	Enhanced Probability
0.95	0.95	0.98
0.95	0.5	0.96
0.95	0.1	0.95
0.5	0.95	0.85
0.5	0.5	0.69
0.5	0.1	0.54
0.1	0.95	0.74
0.1	0.5	0.44
0.1	0.1	0.17

The enhancement of probability leverages supplemental information we can take from the noisy citizen labels, weighting the enhancement according to how well citizens predict that class.

It can only increase the probability the image belongs to the citizen-labeled class.

Semi-Supervised Active Learning Framework



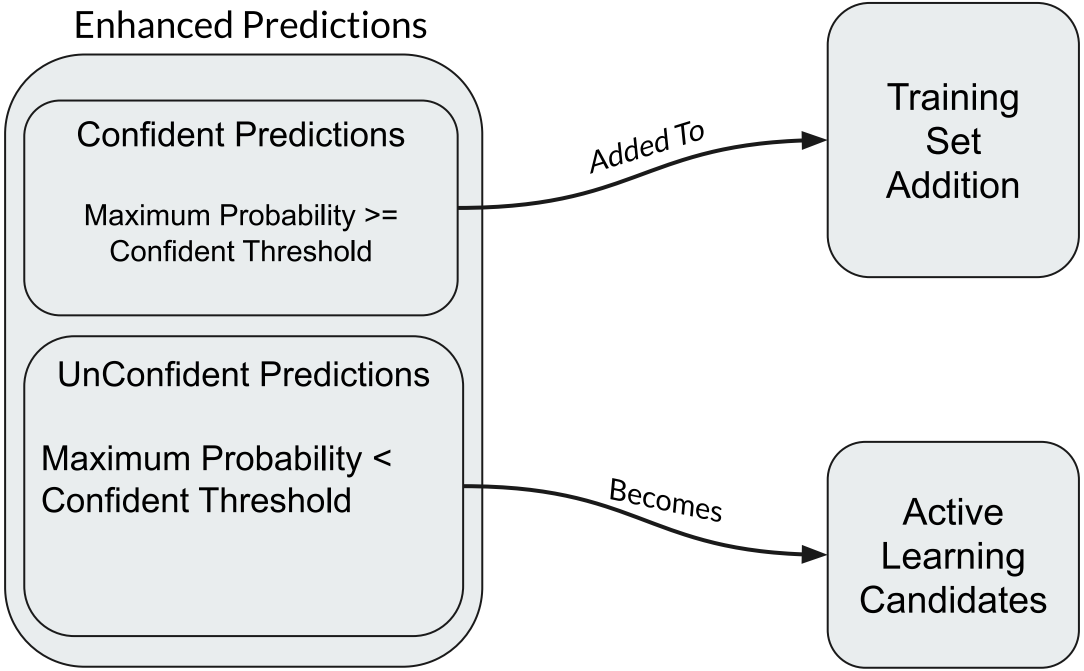
Splitting Predictions and Assigning Pseudolabels

- We set a threshold that defines confident and unconfident predictions (0,1).

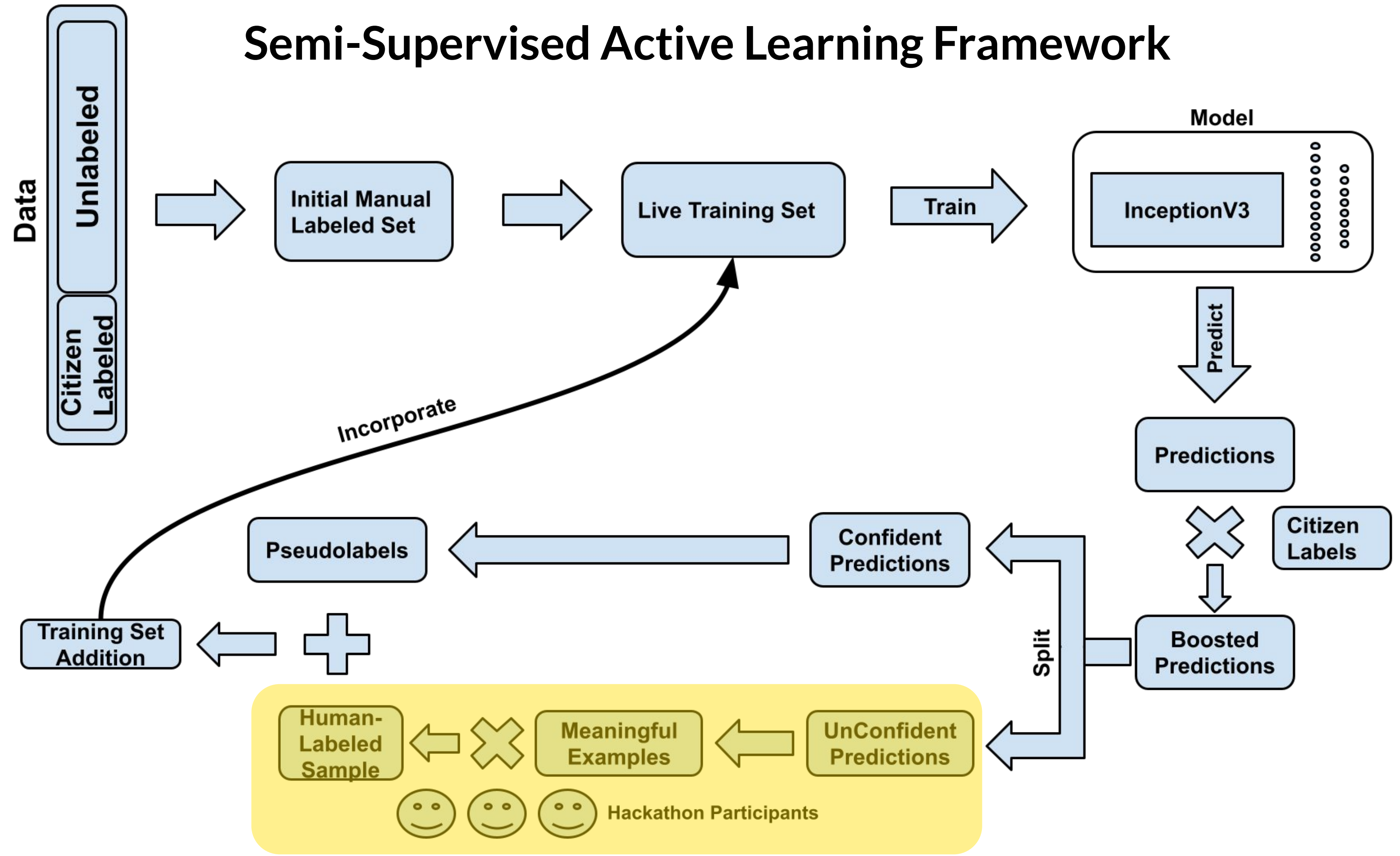
- Predictions with a maximum probability greater than the threshold are trusted and become pseudolabels.

- Predictions which do not pass the threshold do not become pseudolabels.

- Usually the threshold is 0.7-0.95 [41,42].



Semi-Supervised Active Learning Framework



Active Learning - Weighted Entropy Selection

- Selection strategies rank the samples in order of expected usefulness to label.

- Has shown to be much more effective compared to random labeling [43]

- Entropy-Maximization is the go-to selection strategy

- We implore a modified entropy which prioritizes entropy among rare classes via weighting to discover instances which have higher likelihood to belong to a rare class

High Entropy: [0.33, 0.33, 0.33] -> 0.48
Low Entropy: [0.98, 0.01, 0.01] -> 0.05

Entropy:

$$H_j = -\sum_{i=0}^{n-1} (P_j[i] * log(P_j[i]))$$

Weighted Entropy:

$$H_j^w = -\sum_{i=0}^{n-1} (P_j[i] * log(P_j[i]) * log(w[i])^2)$$

H_j - Image j's prediction entropy

H_j^w - Image j's prediction weighted entropy

n - Number of classes

P_j - Predicted probabilities for image j

w - List containing class proportions

Active Learning - Weighted Entropy Example

$$class\_distribution = [0.6, 0.3, 0.1]$$

$$H_j^w = -\sum_{i=0}^{n-1}(P_j[i] * log(P_j[i]) * log(w[i])^2)$$

Example 1

Entropy contained in common classes

$P_j = [0.5, 0.4, 0.1]$

$H_j = 0.94$ (entropy)

$H_j^w = 1.84$ (weighted entropy)

Example 2

Entropy contained in rare classes

$P_j = [0.1, 0.4, 0.5]$

$H_j = 0.94$ (entropy)

$H_j^w = 2.42$ (weighted entropy)

In the two examples, they have the same regular entropy. Ex2 has higher weighted entropy because it has higher entropy in the rare classes. Ex2 would be ranked higher for labeling, because it has higher weighted entropy.

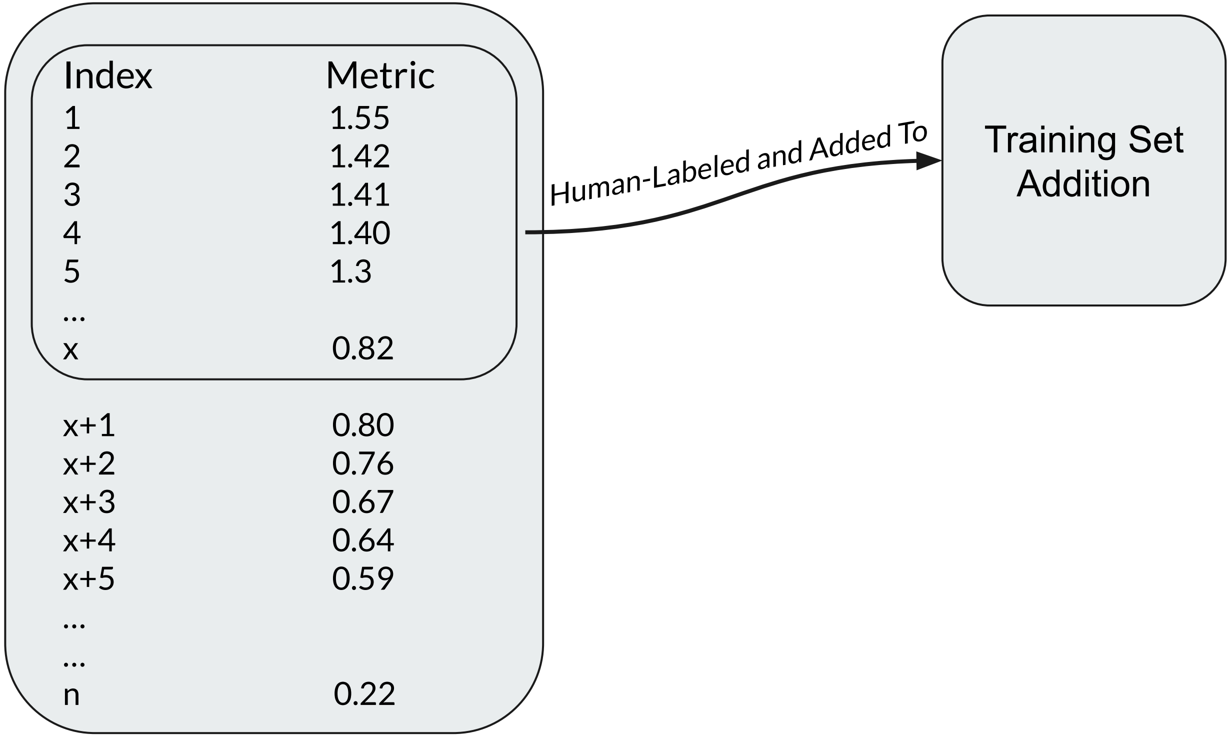
Active Learning Selections

- Calculate selection criteria for each prediction

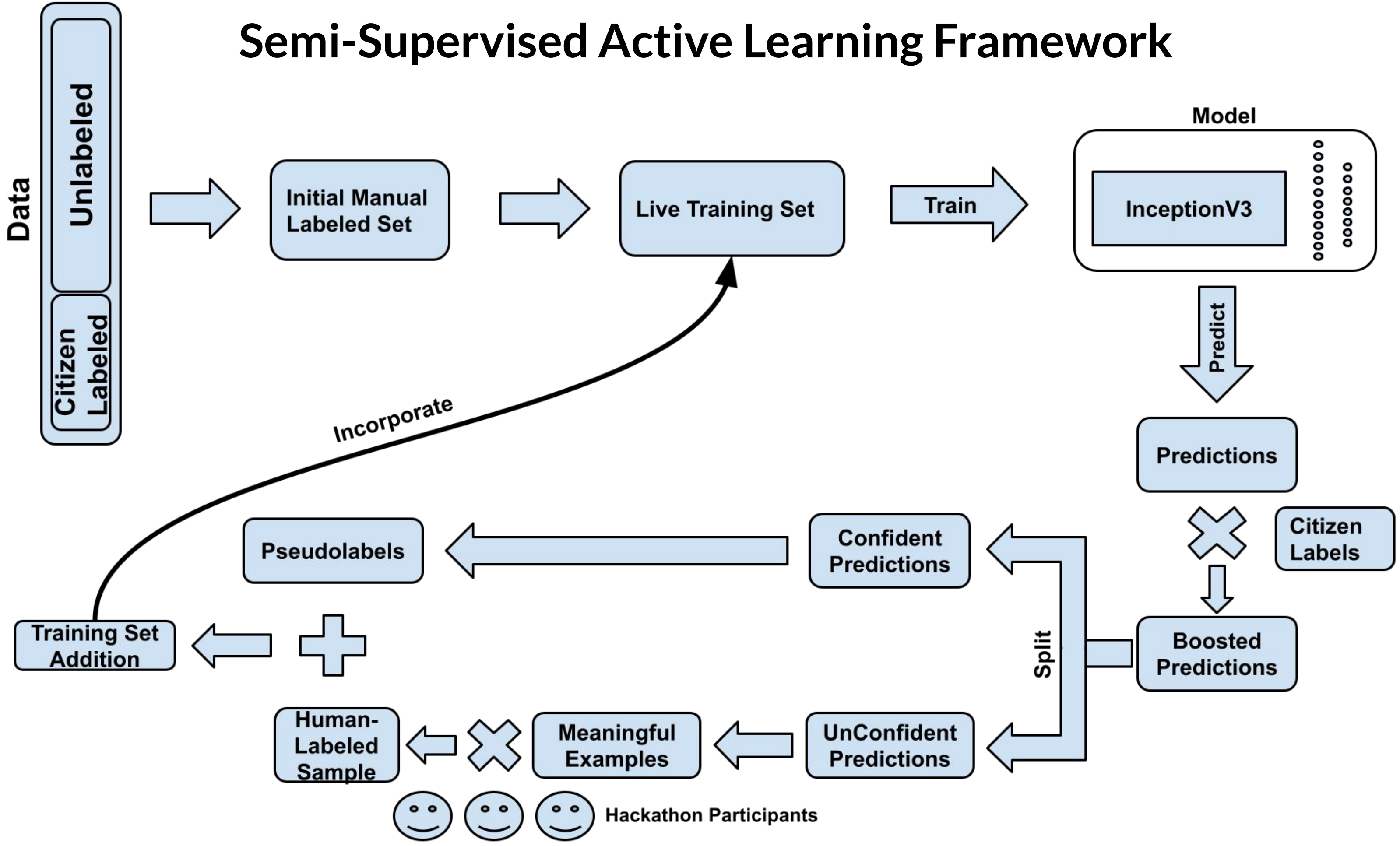
- Sort based on selection criteria

- Have top x samples labeled and added to training set

Unconfident UnEnhanced Predictions



Semi-Supervised Active Learning Framework



GLOBE Experiments

Naive Baseline: Throw out the SSAL Framework and train directly on the ~17,000 noisily labeled images, with a traditional loss function.

SSAL Baseline: Use the SSAL Framework, but without enhancing the predictions with citizen labels and with regular entropy active learning selection.

Our Modified SSAL: Use the SSAL Framework, enhancing predictions with citizen labels, and using the modified, weighted entropy for active learning selection.

Goals:

1) Produce an acceptable model for GLOBE image labeling
2) SSAL perform better than the Naive Baseline
3) Modified SSAL perform better than the SSAL Baseline
4) Enhancing predictions produce more pseudolabels, with comparable accuracy
5) Weighted entropy produces favorable distribution of classes compared to baseline

GLOBE Methods

SSAL Parameters

Initial Training Set Size - 500
Validation Size - 1000
Test Size - 3000

Labeling budget - 2,000 images
Number of iterations - 10
Size of AL sample per iteration - 150
Size of predictions per iteration - 7,500
Citizen Weight - 0.75
Confident Threshold 0.90

Model/Training Parameters

Epochs - 75
Batch Size - 32
Early Stopping Patience - 5
Learning rate - 1e-6
Optimizer - ADAM
Loss - CCE

Base Model - Inception-V3
Label Smoothing - 0.1
Dropout - 0.4
Image Augmentation - None
Weight Classes - True

Meta Experiment Info

Trials - 2
Results averaged

Computing Environment - Discovery
Single P100/V100 GPU
32GB RAM
8 Cores

Deep Learning Environment - Tensorflow

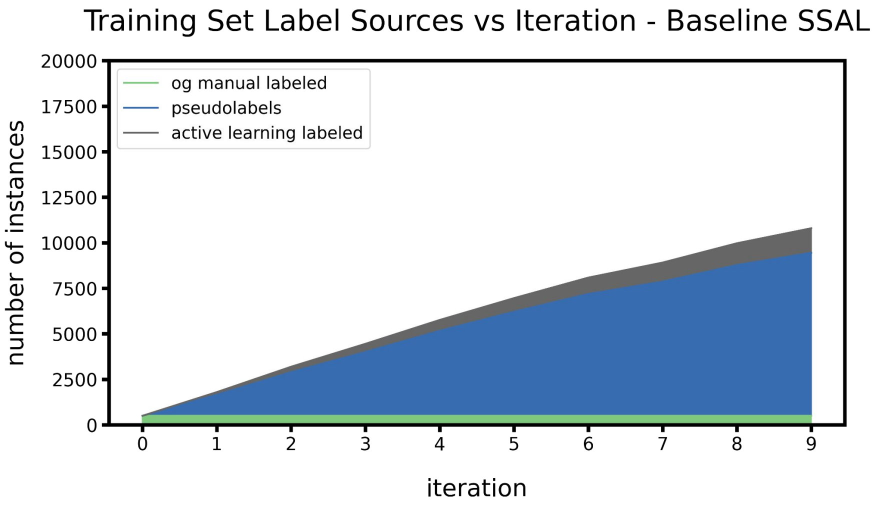
Database - Postgres

GLOBE SSAL Pseudolabeling Results

Baseline

Pseudolabels per iteration - 964 (13% of predictions)

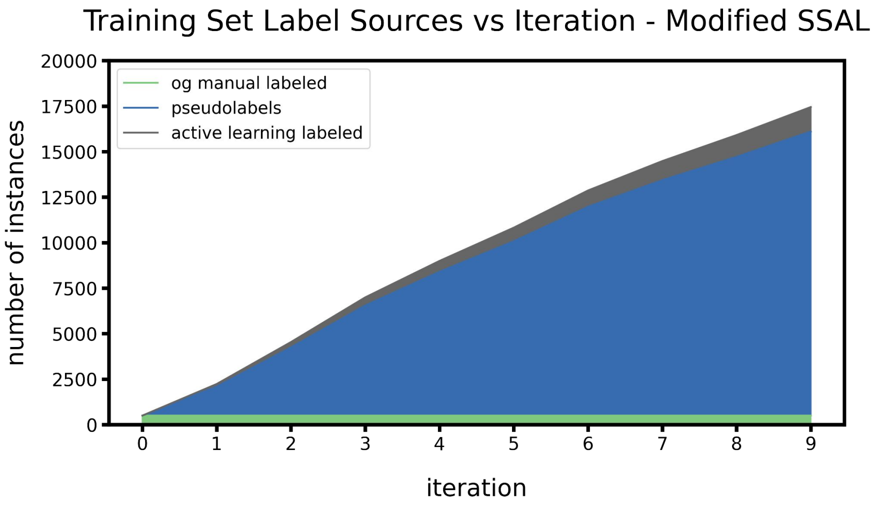
Pseudolabel accuracy - 93%



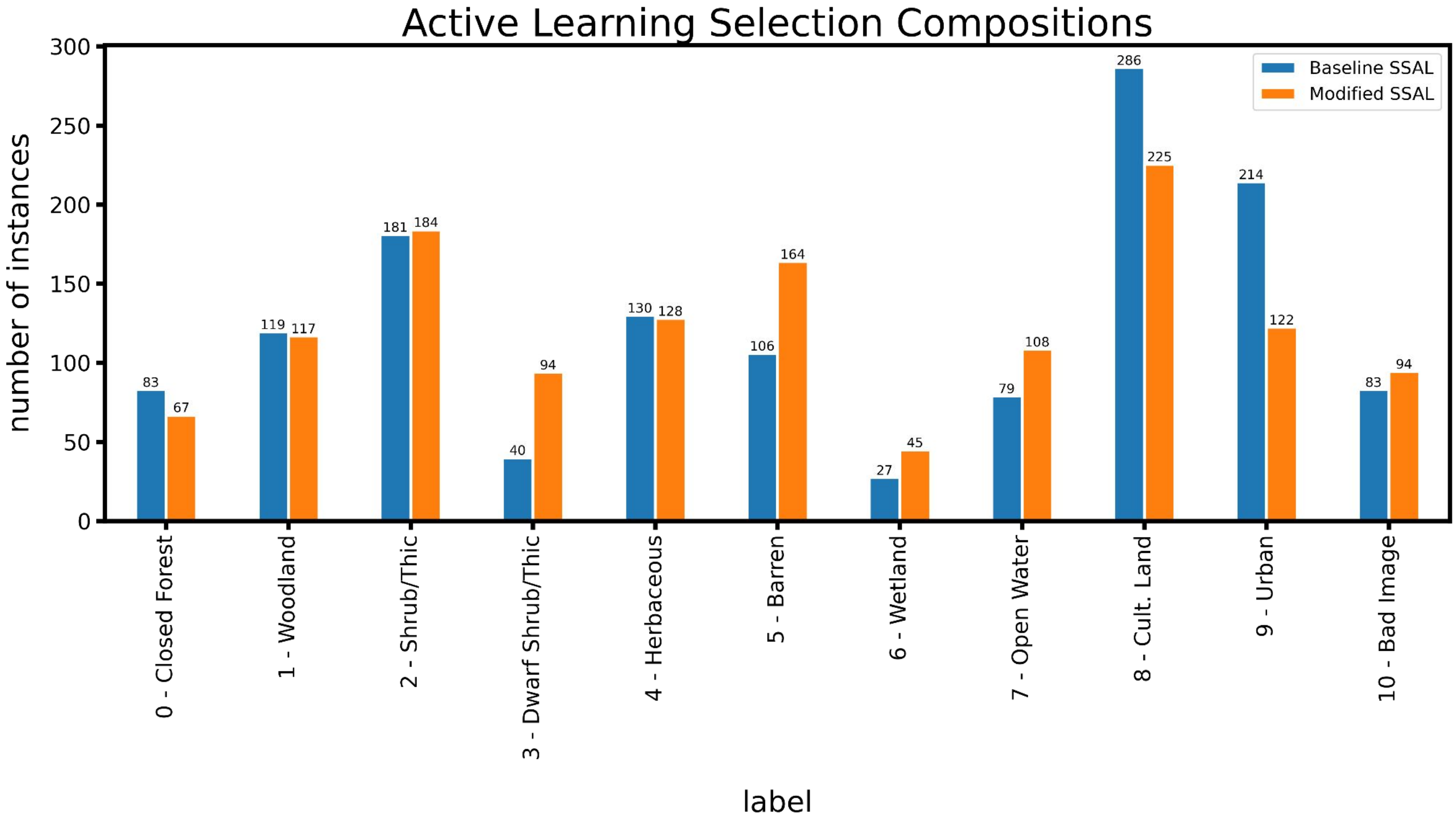
Modified

Pseudolabels per iteration - 1677 (22% of predictions)

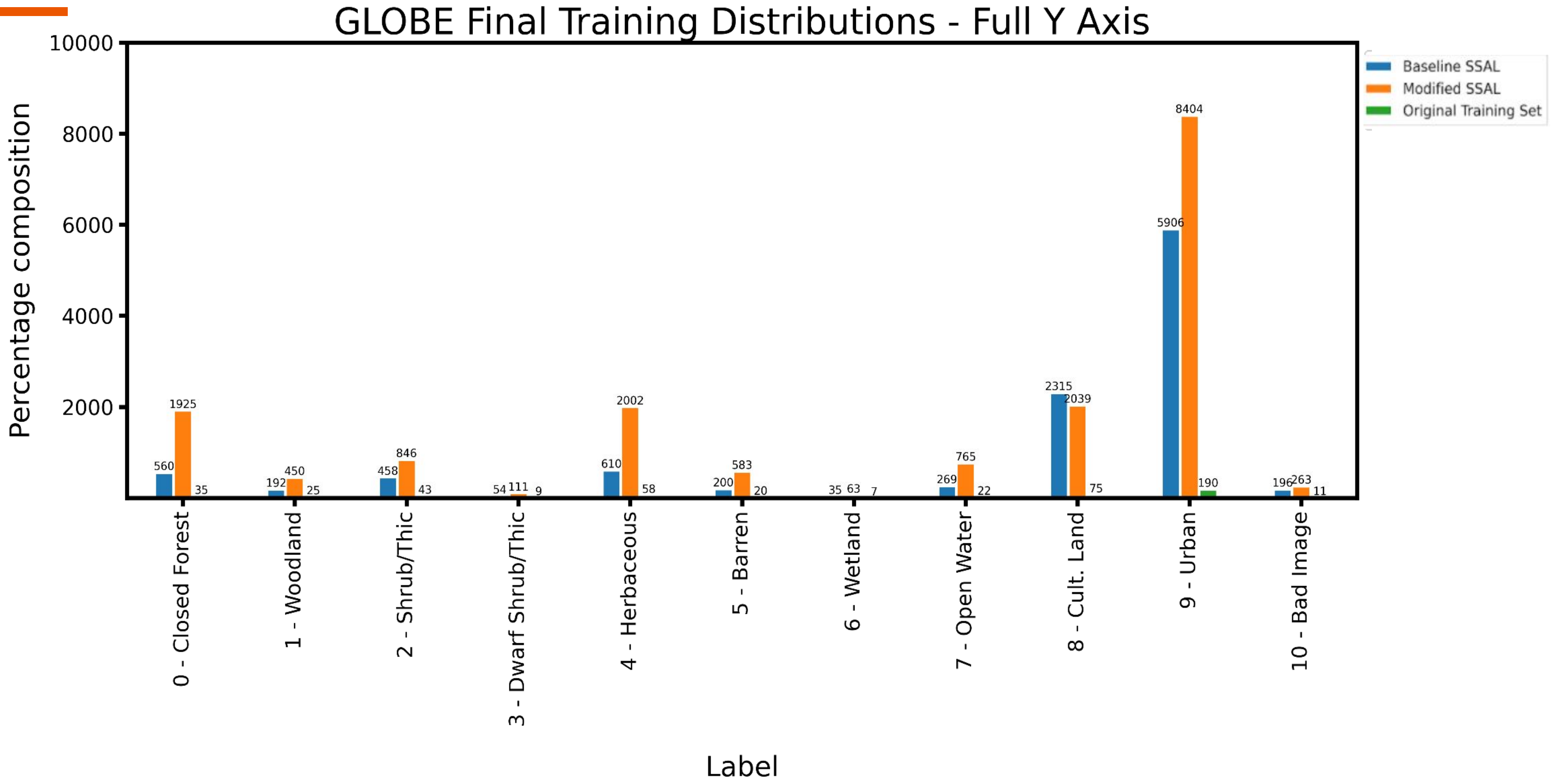
Pseudolabel accuracy - 85%



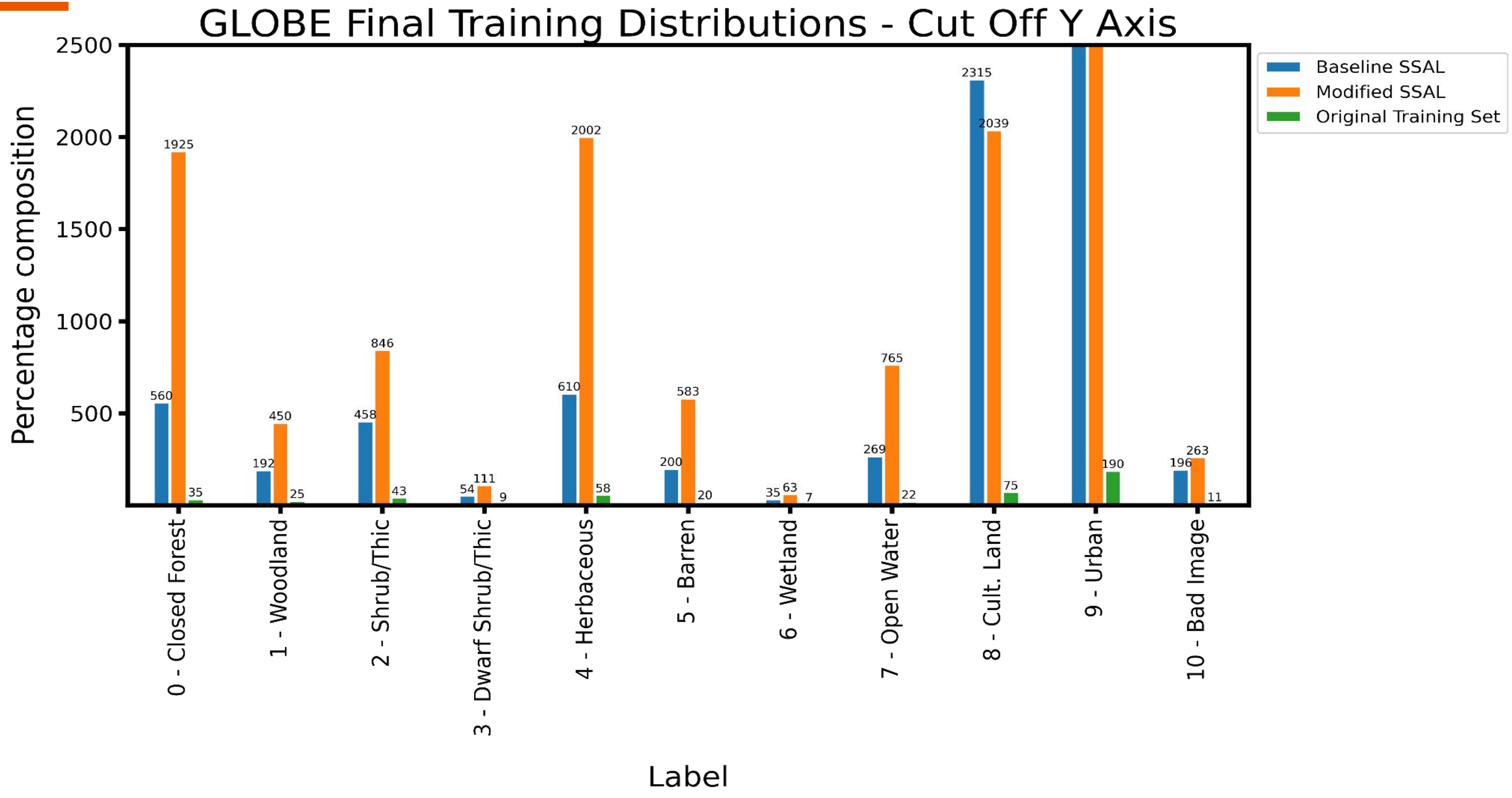
GLOBE SSAL Active Learning Results



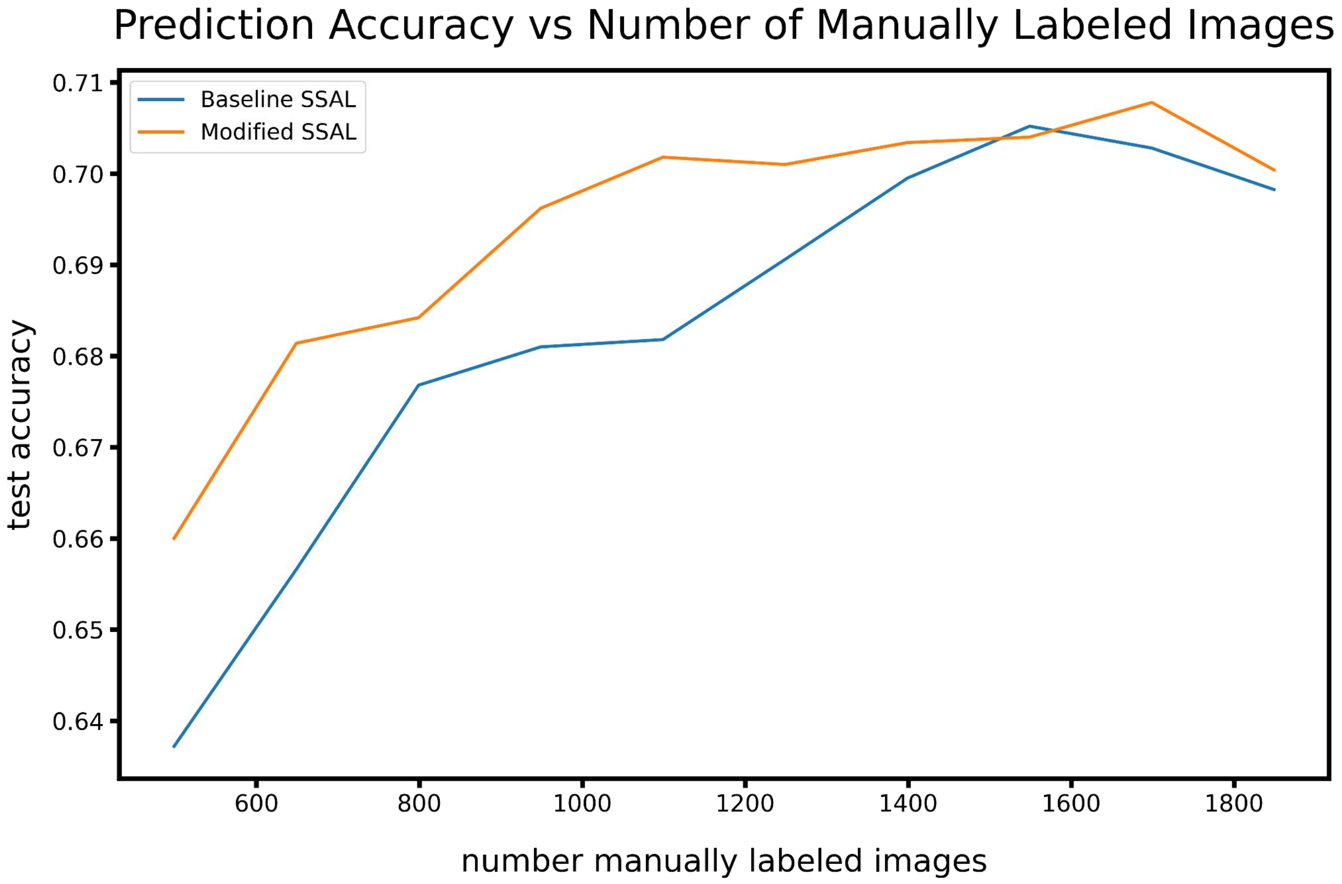
GLOBE SSAL Final Training Set Distribution



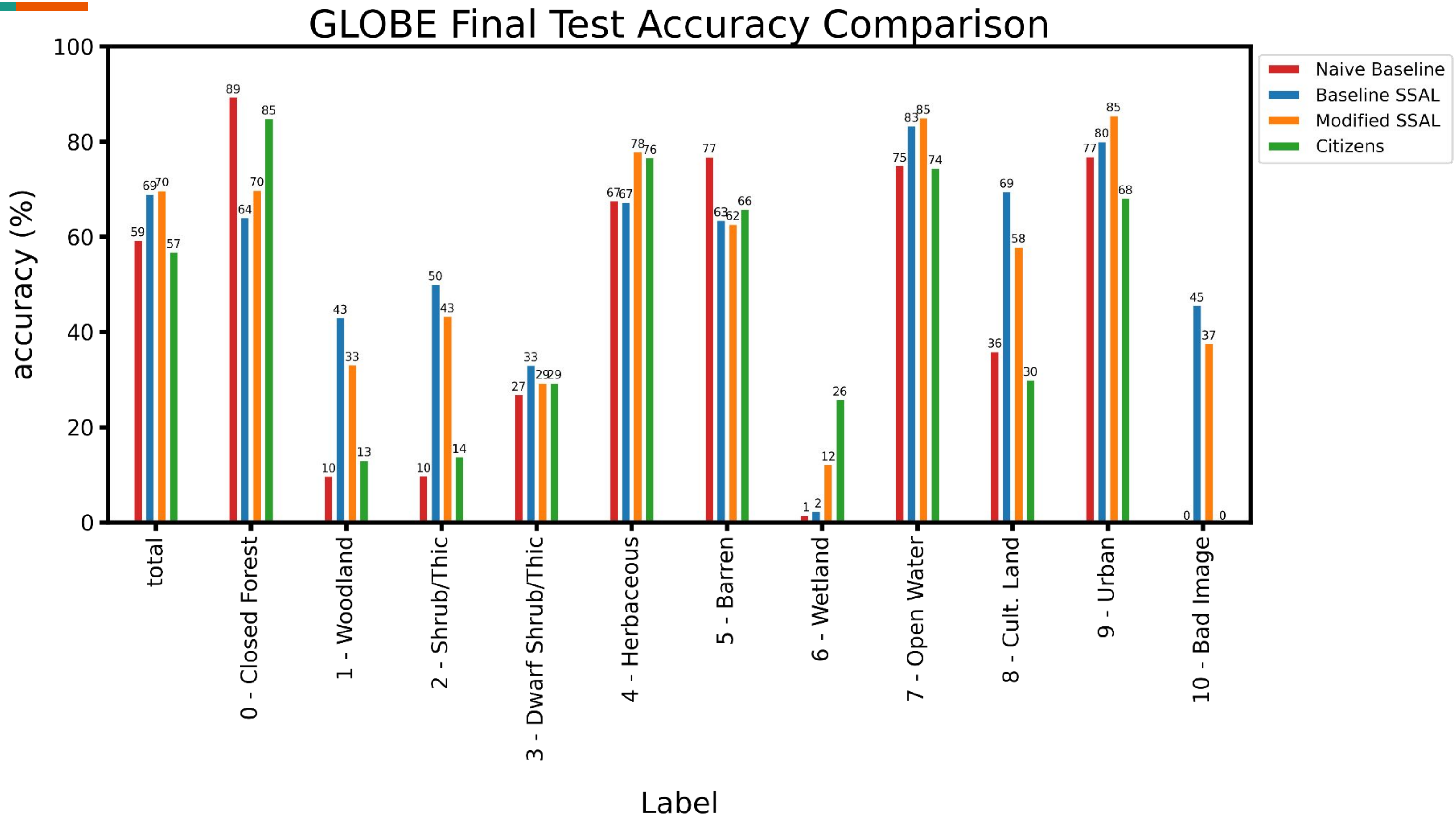
GLOBE SSAL Final Training Set Distribution



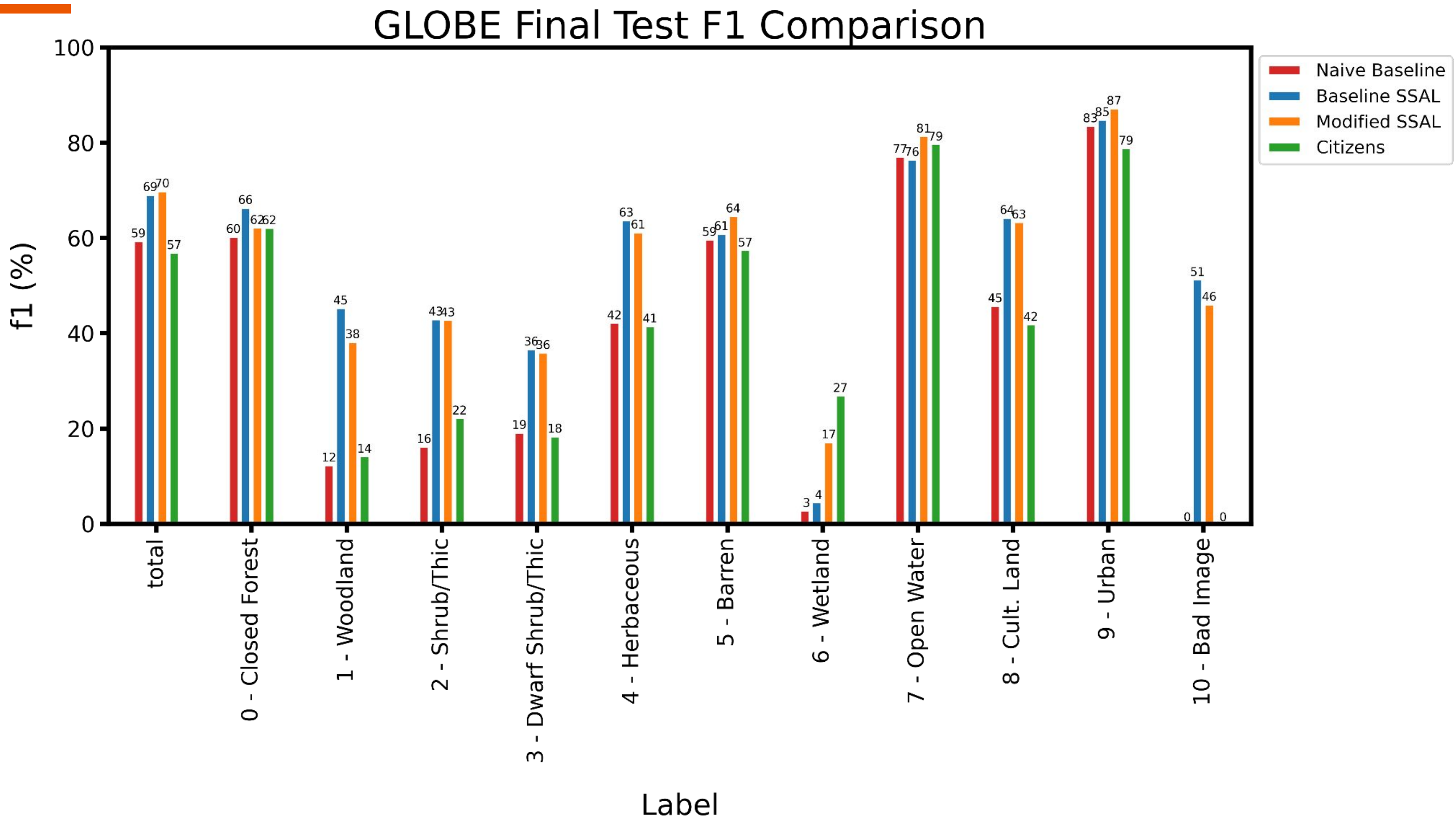
GLOBE SSAL Accuracy vs Number of Labeled Images



GLOBE SSAL Final Test Accuracies



GLOBE SSAL Final Test F1



GLOBE SSAL Confusion Matrices

Baseline SSAL Confusion Matrix

True label \ Predicted label	0 - Closed Forest	1 - Woodland	2 - Shrub/Thic	3 - Dwarf Shrub/Thic	4 - Herbaceous	5 - Barren	6 - Wetland	7 - Open Water	8 - Cult. Land	9 - Urban	10 - Bad Image
0 - Closed Forest	0.64	0.09	0.17	0.00	0.01	0.02	0.00	0.01	0.03	0.03	0.00
1 - Woodland	0.27	0.43	0.14	0.00	0.01	0.03	0.00	0.01	0.06	0.03	0.01
2 - Shrub/Thic	0.07	0.10	0.50	0.02	0.11	0.03	0.01	0.01	0.06	0.08	0.02
3 - Dwarf Shrub/Thic	0.00	0.06	0.28	0.33	0.19	0.11	0.00	0.00	0.02	0.00	0.02
4 - Herbaceous	0.01	0.02	0.09	0.01	0.68	0.02	0.00	0.00	0.12	0.04	0.01
5 - Barren	0.02	0.04	0.06	0.03	0.06	0.64	0.00	0.06	0.07	0.04	0.01
6 - Wetland	0.08	0.11	0.03	0.00	0.16	0.00	0.03	0.45	0.05	0.11	0.00
7 - Open Water	0.00	0.02	0.01	0.00	0.01	0.03	0.00	0.84	0.01	0.06	0.02
8 - Cult. Land	0.00	0.02	0.03	0.00	0.15	0.03	0.00	0.01	0.70	0.06	0.00
9 - Urban	0.00	0.01	0.02	0.00	0.01	0.02	0.00	0.01	0.11	0.80	0.01
10 - Bad Image	0.04	0.01	0.06	0.00	0.04	0.03	0.00	0.01	0.12	0.23	0.46

Predicted label
accuracy=0.6932; misclass=0.3068

Modified SSAL Confusion Matrix

True label \ Predicted label	0 - Closed Forest	1 - Woodland	2 - Shrub/Thic	3 - Dwarf Shrub/Thic	4 - Herbaceous	5 - Barren	6 - Wetland	7 - Open Water	8 - Cult. Land	9 - Urban	10 - Bad Image
0 - Closed Forest	0.70	0.11	0.10	0.00	0.01	0.02	0.00	0.00	0.01	0.03	0.00
1 - Woodland	0.38	0.33	0.09	0.00	0.08	0.02	0.01	0.00	0.04	0.05	0.01
2 - Shrub/Thic	0.22	0.03	0.44	0.02	0.16	0.03	0.00	0.00	0.03	0.07	0.01
3 - Dwarf Shrub/Thic	0.00	0.00	0.20	0.30	0.37	0.11	0.00	0.00	0.00	0.02	0.00
4 - Herbaceous	0.02	0.01	0.07	0.00	0.78	0.02	0.00	0.00	0.06	0.04	0.00
5 - Barren	0.00	0.05	0.04	0.02	0.07	0.63	0.00	0.05	0.05	0.07	0.02
6 - Wetland	0.14	0.03	0.06	0.00	0.14	0.03	0.11	0.34	0.09	0.06	0.00
7 - Open Water	0.02	0.02	0.01	0.00	0.00	0.04	0.01	0.85	0.00	0.03	0.01
8 - Cult. Land	0.01	0.01	0.02	0.00	0.26	0.01	0.00	0.00	0.58	0.10	0.00
9 - Urban	0.01	0.01	0.02	0.00	0.02	0.01	0.00	0.01	0.06	0.86	0.01
10 - Bad Image	0.11	0.00	0.03	0.01	0.11	0.02	0.00	0.02	0.06	0.27	0.38

Predicted label
accuracy=0.7010; misclass=0.2990

GLOBE Conclusion

- SSAL outperforms Naive Baseline

- The Baseline and Modified SSAL strategies have same final performance with Modified SSAL getting there faster.

- Weighted entropy was able to find more rare-class instances during active learning, but no improvement in accuracy.

- Enhancing predictions with citizen labels nearly doubled pseudolabel production

- Overall, were able to train a model which outperforms citizens, but not ideal.

	Test Accuracy (%)	Test Accuracy @ 1000 Human Labels (%)	Top2 Test Accuracy (%)
Naive Baseline	59	N/A	63
Baseline SSAL	69	68	85
Modified SSAL	70	70	84
Citizen Scientists	57	N/A	57

Intel Image Scene Dataset

- Contains 6 classes, each with ~3,000 images.

- Covers a similar domain as GLOBE

- Known to be clean, reliable, balanced

- Ran experiments with similar, simulated conditions as GLOBE dataset

- Comes pre-labeled, so less work for me

0 - Buildings

3 - Mountain

1 - Forest

4 - Sea

2 - Glacier

[2]

5 - Street

Intel Image Scene Dataset Experiments

- Ran four experiments with varying degrees of imbalance and noise, starting with no noise and no imbalance.

- Inject noise by switching a percentage of class's labels to random value

- Inject imbalance by making a portion of class's images unavailable.

- Ran Naive Baseline, Baseline SSAL, Modified SSAL trials.

- Also ran trials with noise-robust loss functions, MAE and Truncated-CCE loss [4]

Experiment 1 - No noise, No imbalance

Noise - [0,0,0,0,0,0]

Imbalance - [1,1,1,1,1,1]

Experiment 2 - Low noise, imbalance

Noise - [0.25, 0.35, 0.30, 0.20, 0.25, 0.35]

Imbalance - [1, 0.8, 0.7, 0.6, 0.4, 0.2]

Experiment 3 - Modest noise, imbalance

Noise - [0.45, 0.55, 0.40, 0.50, 0.35, 0.55]

Imbalance - [1, 0.6, 0.4, 0.35, 0.2, 0.1]

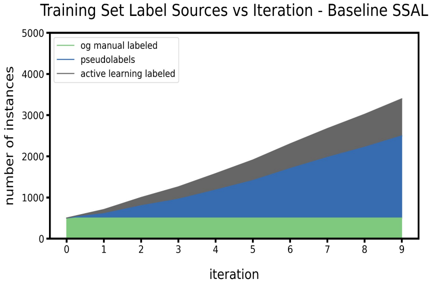
Experiment 4 - Extreme noise, imbalance

Noise - [0.45, 0.55, 0.50, 0.70, 0.45, 0.65]

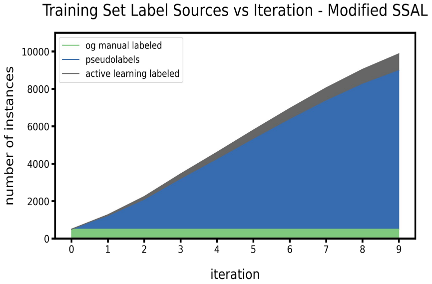
Imbalance - [1.0, 0.6, 0.30, 0.15, 0.1 ,0.05]

Intel Experiment's Pseudolabeling Results

Experiment 1

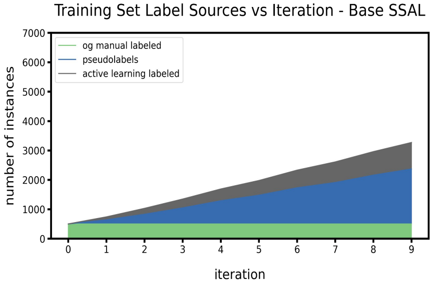


pseudolabels/iteration - 227

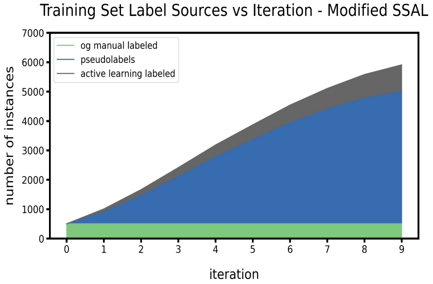


pseudolabels/iteration - 904

Experiment 2

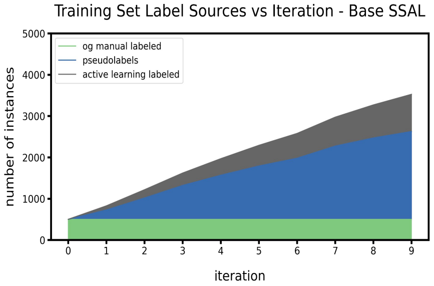


pseudolabels/iteration - 208

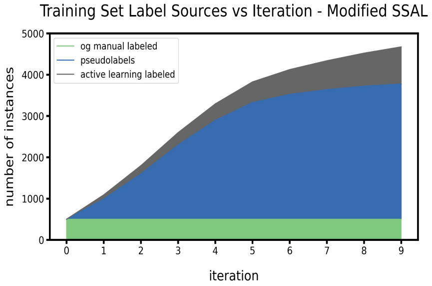


pseudolabels/iteration - 466

Experiment 3

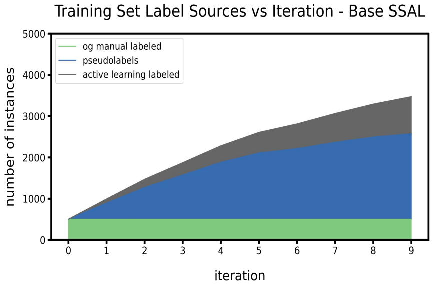


pseudolabels/iteration - 227

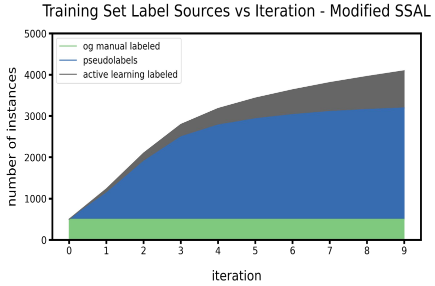


pseudolabels/iteration - 323

Experiment 4



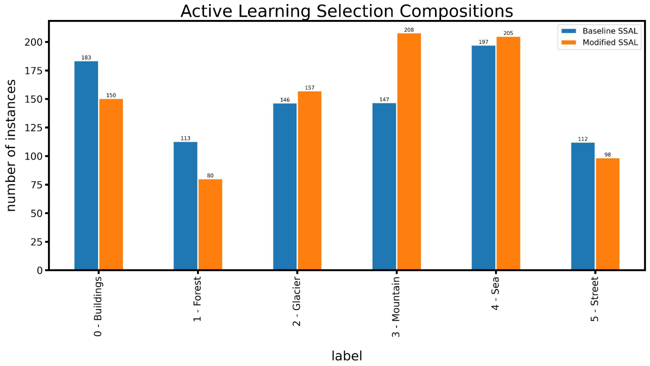
pseudolabels/iteration - 213



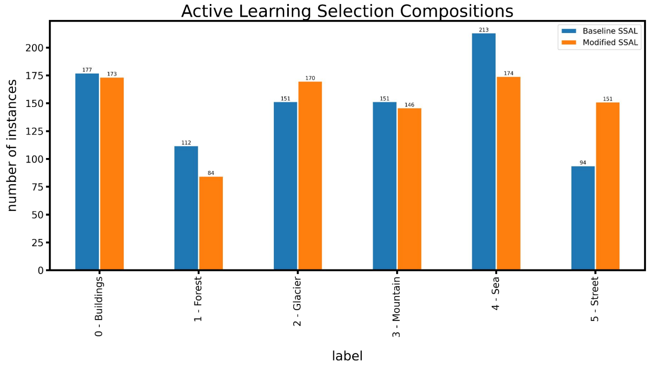
pseudolabels/iteration - 271

Intel Experiment's Active Learning Results

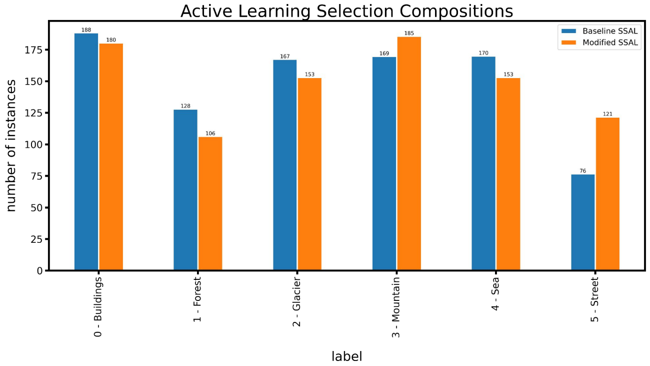
Experiment 1



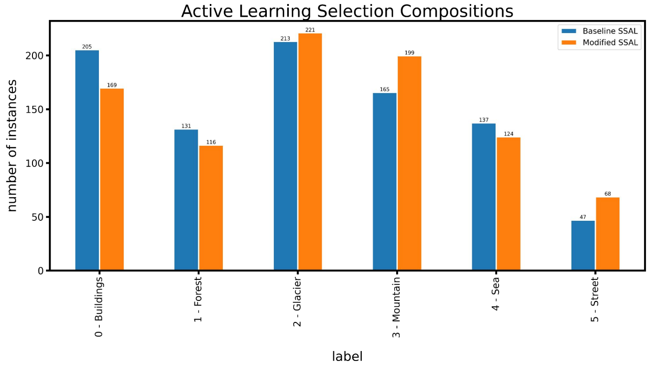
Experiment 2



Experiment 3

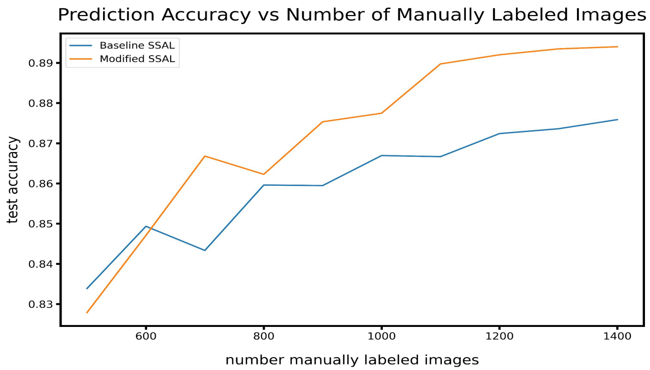


Experiment 4

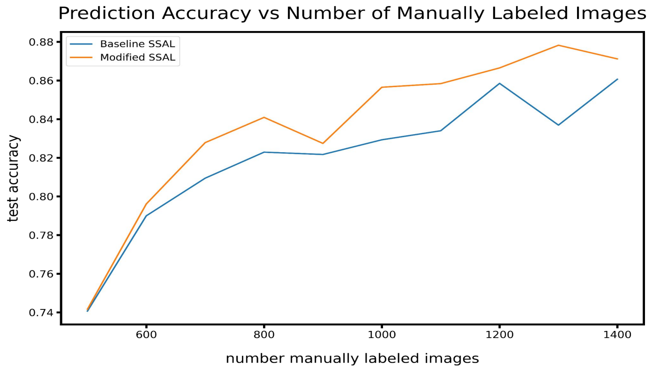


Intel SSAL Accuracy vs Num Labeled Images

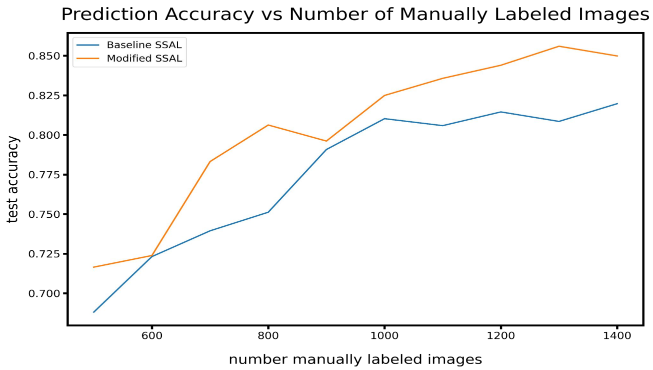
Experiment 1



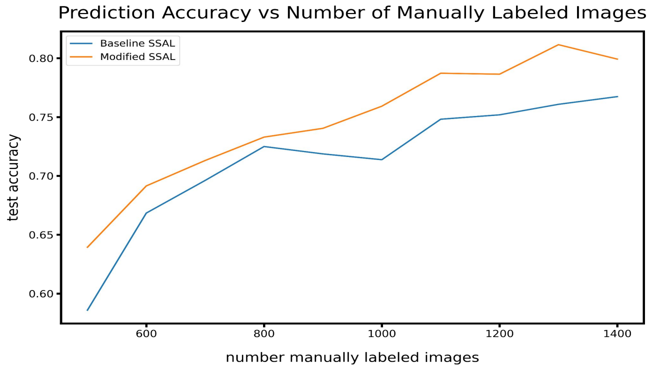
Experiment 2



Experiment 3

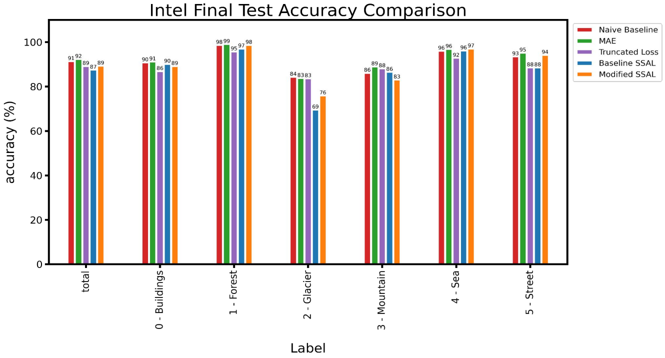


Experiment 4

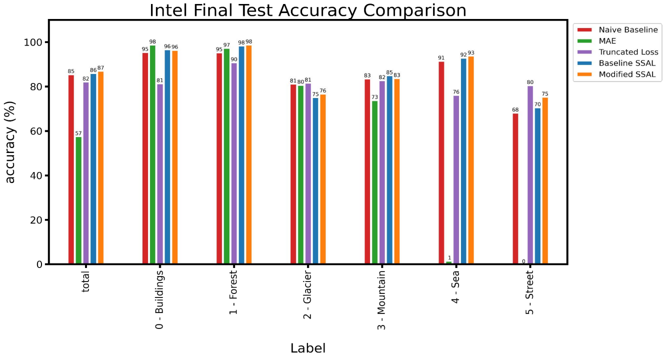


Intel SSAL Accuracy vs Num Labeled Images

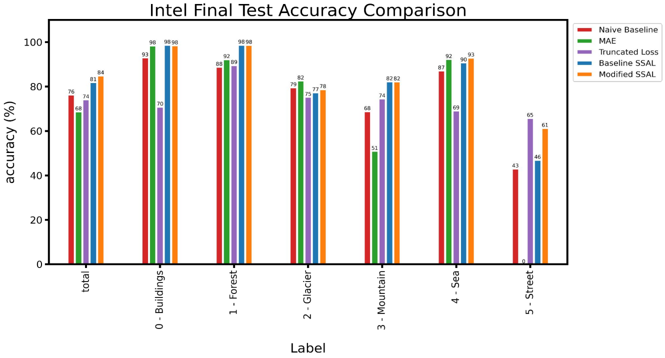
Experiment 1



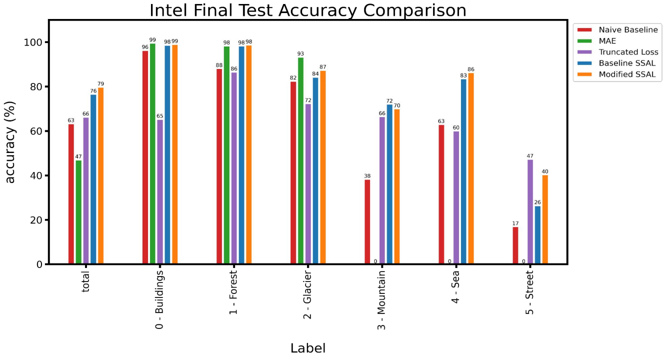
Experiment 2



Experiment 3



Experiment 4



Intel Conclusion

- With more noise and imbalance, SSAL comparative performance improves

- Modified SSAL slightly outperforms Baseline SSAL with more noise and imbalance

- Weighted entropy was able to find slightly more instances of rare classes in general

- Relative increase in pseudolabel production decreases with more noise

- Modified SSAL very slightly outperformed Baseline SSAL in rare classes

	Experiment 1 Accuracy	Experiment 2 Accuracy	Experiment 3 Accuracy	Experiment 4 Accuracy
Naive Baseline	91	85	76	63
MAE	92	57	68	47
Truncated Loss	89	82	74	66
Baseline SSAL	87	86	81	76
Modified SSAL	89	87	84	79

Future Work

- More related work experiments.

- Hackathon in early 2022.

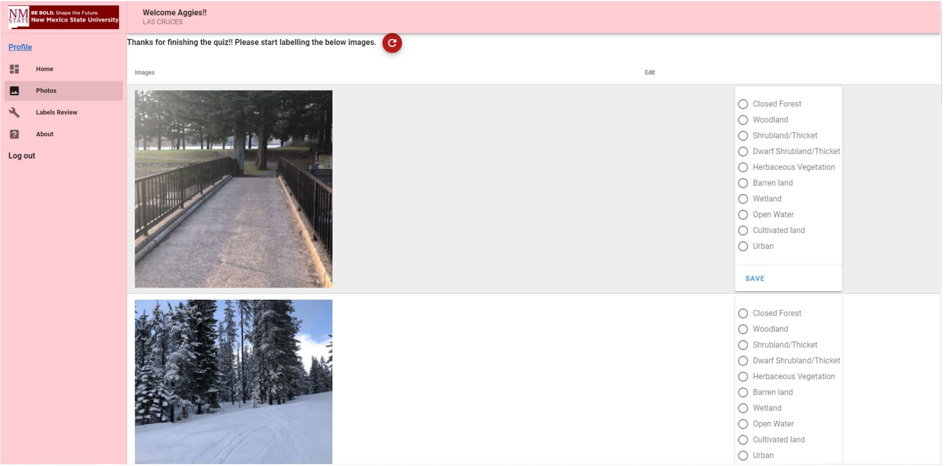
- Derive MUC code from all N/E/S/W images, rather than a single one.

- Prediction of composition of classes in image. Semantic Segmentation?

- Address poor images

- Role of model to NASA

Hackathon interface, by Preeti Maurya



Semantic segmentation for land cover [3]

Citations

[1] Global Learning and Observations to Benefit the Environment (GLOBE) Program, January 27th, 2021

[2] Intel image classification challenge. https://www.kaggle.com/puneet6060/ intel-image-classification.\

[3] S. Xu, S. Zhang, J. Zeng, T. Li, Q. Guo and S. Jin, "A Framework for Land Use Scenes Classification Based on Landscape Photos," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6124-6141, 2020, doi: 10.1109/JSTARS.2020.3028158

[4] Zhang, Zhilu, and Mert R. Sabuncu. "Generalized cross entropy loss for training deep neural networks with noisy labels." *32nd Conference on Neural Information Processing Systems (NeurIPS)*. 2018.

[5] Barry-Straume, Jostein; Tschannen, Adam; Engels, Daniel W.; and Fine, Edward (2018) "An Evaluation of Training Size Impact on Validation Accuracy for Optimized Convolutional Neural Networks," SMU Data Science Review: Vol. 1 : No. 4 , Article 12.

[6] Althnian, A.; AlSaeed, D.; Al-Baity, H.; Samha, A.; Dris, A.B.; Alzakari, N.; Abou Elwafa, A.; Kurdi, H. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. Appl. Sci. 2021, 11, 796. https://doi.org/10.3390/app11020796

[7] Rukundo, Olivier. (2021). EFFECTS OF IMAGE SIZE ON DEEP LEARNING.

[8] Silva, Wilson & Schwerz de Lucena, Diogo. (2018). Concrete Cracks Detection Based on Deep Learning Image Classification. Proceedings. 2. 5387. 10.3390/ICEM18-05387.

[9] Algan, Gorkem and Ilkay Ulusoy. "Label Noise Types and Their Effects on Deep Learning." *ArXiv* abs/2003.10471 (2020): n. pag.

[10] Song, Hwanjun, et al. "Learning from noisy labels with deep neural networks: A survey." *arXiv preprint arXiv:2007.08199* (2020).

[11] Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. J Big Data 6, 27 (2019). https://doi.org/10.1186/s40537-019-0192-5

[12] Buda, Mateusz, Atsuto Maki, and Maciej A. Mazurowski. "A systematic study of the class imbalance problem in convolutional neural networks." *Neural Networks* 106 (2018): 249-259.

[13] Johnson, J.M., Khoshgoftaar, T.M. The Effects of Data Sampling with Deep Learning and Highly Imbalanced Big Data. *Inf Syst Front* 22, 1113–1131 (2020). https://doi.org/10.1007/s10796-020-10022-7

[14] Thabtah, Fadi & Hammoud, Suhel & Kamalov, Firuz & Gonsalvesv, Amanda. (2019). Data Imbalance in Classification: Experimental Evaluation. Information Sciences. 513. 10.1016/j.ins.2019.11.004.

[15] Castillo Lamas, Alberto & Tabik, Siham & Cruz, Policarpo & Montes, Rosana & Martinez, Alvaro & Cruz, Teresa & Herrera, Francisco. (2021). MonuMAI: Dataset, deep learning pipeline and citizen science based app for monumental heritage taxonomy and classification. Neurocomputing. 420. 266-280. 10.1016/j.neucom.2020.09.041.

[16] E. Maggiori, Y. Tarabalka, G. Charpiat and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645-657, Feb. 2017, doi: 10.1109/TGRS.2016.2612821.

[17] Buscombe D, Ritchie AC. Landscape Classification with Deep Neural Networks. *Geosciences*. 2018; 8(7):244. https://doi.org/10.3390/geosciences8070244

[18] Genbatu Ge, Zhongjie Shi, Yuanjun Zhu, Xiaohui Yang, Yuguang Hao, Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms, Global Ecology and Conservation, Volume 22, 2020, e00971, ISSN 2351-9894, https://doi.org/10.1016/j.gecco.2020.e00971.

Citations

[19] Perez, Luis, and Jason Wang. "The effectiveness of data augmentation in image classification using deep learning." arXiv preprint arXiv:1712.04621 (2017).

[20] Goodfellow, I. (n.d.). Regularization for Deep Learning. Lecture.

[21] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

[22] Wei Zhao , "Research on the deep learning of the small sample data based on transfer learning", AIP Conference Proceedings 1864, 020018 (2017) https://doi.org/10.1063/1.4992835

[23] Hussain M., Bird J.J., Faria D.R. (2019) A Study on CNN Transfer Learning for Image Classification. In: Lotfi A., Bouchachia H., Gegov A., Langensiepen C., McGinnity M. (eds) Advances in Computational Intelligence Systems. UKCI 2018. Advances in Intelligent Systems and Computing, vol 840. Springer, Cham. https://doi.org/10.1007/978-3-319-97982-3_16

[24] H. Ranganathan, H. Venkateswara, S. Chakraborty and S. Panchanathan, "Deep active learning for image classification," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3934-3938, doi: 10.1109/ICIP.2017.8297020.

[25] Tanaka, Daiki, et al. "Joint optimization framework for learning with noisy labels." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[26] Veit, Andreas, et al. "Learning from noisy large-scale datasets with minimal supervision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[27] Han, Jiangfan, Ping Luo, and Xiaogang Wang. "Deep self-learning from noisy labels." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[28] Lee, Dong-Hyun. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks." Workshop on challenges in representation learning, ICML. Vol. 3. No. 2. 2013.

[29] Lukasik, Michal, et al. "Does label smoothing mitigate label noise?." International Conference on Machine Learning. PMLR, 2020.

[30] Q. Dong, S. Gong and X. Zhu, "Imbalanced Deep Learning by Minority Class Incremental Rectification," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 6, pp. 1367-1381, 1 June 2019, doi: 10.1109/TPAMI.2018.2832629

[31] Ando, Shin, and Chun Yuan Huang. "Deep over-sampling framework for classifying imbalanced data." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2017.

[32] Lin, Christopher H., Mausam Mausam, and Daniel S. Weld. "Active Learning with Unbalanced Classes and Example-Generation Queries." Sixth AAAI Conference on Human Computation and Crowdsourcing. 2018.

[33] Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. J Big Data 6, 27 (2019). https://doi.org/10.1186/s40537-019-0192-5

[34] Lin T-Y, Goyal P, Girshick RB, He K, Dollár P. Focal loss for dense object detection. In: IEEE international conference on computer vision (ICCV). vol. 2017. 2017. p. 2999–3007.

[35] Ren, Mengye, et al. "Learning to reweight examples for robust deep learning." International Conference on Machine Learning. PMLR, 2018..

[36] Koziarski, Michał, Michał Woźniak, and Bartosz Krawczyk. "Combined cleaning and resampling algorithm for multi-class imbalanced data with label noise." Knowledge-Based Systems 204 (2020): 106223.

[37] Zhong, Yaoyao, et al. "Unequal-training for deep face recognition with long-tailed noisy data." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[38] Chen, Baiyun, et al. "RSMOTE: A self-adaptive robust SMOTE for imbalanced problems with label noise." Information Sciences 553 (2021): 397-428.

[39] Chen, D., Liu, S., Kingsbury, P. et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. npj Digit. Med. 2, 43 (2019). https://doi.org/10.1038/s41746-019-0122-0

[40] Kang, Y.; Cho, N.; Yoon, J.; Park, S.; Kim, J. Transfer Learning of a Deep Learning Model for Exploring Tourists' Urban Image Using Geotagged Photos. ISPRS Int. J. Geo-Inf. 2021, 10, 137. https://doi.org/10.3390/ijgi10030137

[41] Sohn, Kihyuk, et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." arXiv preprint arXiv:2001.07685 (2020).

[42] Rizve, Mamshad Nayeem, et al. "In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning." arXiv preprint arXiv:2101.06329 (2021).

[43] Ren, Pengzhen, et al. "A survey of deep active learning." arXiv preprint arXiv:2009.00236 (2020).