CISSS-Es: Curriculum-Including Semi-Supervised Sub-Ensembles for Plant Phenotyping

Anonymous ECCV 2024 Submission

Paper ID #0026

1 Introduction

Remote sensing data campaigns are at the heart of monitoring invasive species and the management of counteractions. However, challenges arise from limited species data, high degree of variation as plants evolve, changing environments and lighting conditions, and sensor-induced variations [6]. Additionally, label quantity and quality required for robust Deep Learning models are difficult to acquire, as only limited samples with educated consistent labeling can be provided [1, 7]. Remote locations with limited bandwidths and tight timeframes further limit recruiting large computational resources or qualified experts. A pipeline to combine self-supervised learning with Sub-Ensembles for out-of-distribution (OOD) detection and reducing the labeling effort [5, 9] by selecting high-entropy-samples [8] is proposed in this research. The pipeline is demonstrated with a resnet-18 model [2, 10] foundation model (ImageNet), and is trainable on consumer-grade hardware overnight and therefore applicable to field-use.

2 Materials and Methods

Sub-Ensembles [9] can be used as an approximation to Deep Ensembles [4], which quantify uncertainty through the cross-entropy of class indices:

$$H(\mathbf{x}) = -\sum_{c \in C} p_c(y_c|\mathbf{x}) \ log \ p_c(y_c|\mathbf{x}) \tag{1}$$

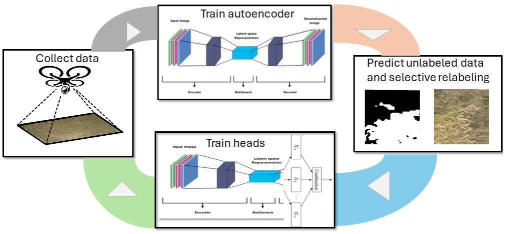

Fig. 1: Process pipeline. Initally, a set of labeled examples is required. Raw data from a new site is used to train an autoencoder (AE). The backbone and bottleneck of the AE are used to train multiple classification heads as a Sub-Ensemble, which is used to predict unseen images, sort by entropy and add to the dataset for the next site.

Table 1: Image Classification Accuracy across Sites and Ablations

				Site 2	Site 3	Site 4
			Growth State	Flowering	Vegetative	Vegetative
Selection	**Backbone**	**Heads**	**True Class**	75.55%	11.87%	50.51%
full	ae	single		57.86%	34.39%	**80.61%**
	ae denoise	single		47.96%	33.87%	49.53 %
100 random	ae	single		52.06%	16.63%	51.38%
	ae denoise	single		26.83%	78.69%	49.53%
100 random	ae	ensmb. 10		74.55%	9.52%	50.51%
	ae denoise	ensmb. 10		**74.62%**	11.96%	72.16%
100 entropy-based	ae	ensmb. 10		74.55%	37.10%	49.49%
	ae denoise	ensmb. 10		**74.61%**	**90.55%**	49.49%

where $p(y_c|\mathbf{x})$ is estimated by the mean of all heads. Training an AE on raw images is proposed as a pretext-task due to limited labeled data. Augmentations such as Masked-Image-Modeling have been shown to boost OOD performance [5]. The pipeline of Fig. 1 is demonstrated on datasets from four sites and image classification is used to identify African lovegrass on orthomosaic patches. True labels are generated using overlaid and upsampled multispectral orthomosaics [3]. AEs are trained on random 32x32 crops from 2000 256x256 patches from high-resolution imagery.

Ablations replace the sampling by entropy with random sampling. AE pretraining is staged into denoising AE or standard AE. As a baseline a resnet with a single head is trained and with limited samples or with the full labeled dataset of the previous sites.

3 Results

Tab. 1 shows that Sub-Ensembles may yield benefits when sufficient information is included. When data is scarce, such as between sites 2 and 3 with few samples, increase is negligible. Using AE backbones does not yield any increases, however, denoising AEs do. Note the large differences in true classes between sites, as well as the changing growth state from site 2 to 3. An entire cycle can be trained in less than 12 hours, requiring under 4 GB GPU memory.

4 Conclusion

Results show that sampling by entropy can improve the classification performance and that denoising AEs improve the performance over only using labeled samples. Denoising AEs capture global image properties as opposed to local details and the primary appearance difference between flowering and vegetative ALG are fewer fine flower-strands. The inconsistent, yet promising, results in Tab. 1 highlight that more research is necessary with the potential of trunk network improvements yielding the required consistency [9]. Future research will address the combination of pipeline elements, such as sophisticated AE architectures and training schemes and the impact of sampling. Also larger scale and distributed scenarios (via model merging techniques) will be explored.

References

1. Geiger, R.S., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J., Tang, R.: "Garbage in, garbage out" revisited: What do machine learning application papers report about human-labeled training data? Quantitative Science Studies **2**(3), 795–827 (Nov 2021). https://doi.org/10.1162/qss_a_00144 1
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016), https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html 1
3. Keerthinathan, P., Amarasingam, N., Kelly, J.E., Mandel, N., Dehaan, R.L., Zheng, L., Hamilton, G., Gonzalez, F.: African Lovegrass Segmentation with Artificial Intelligence Using UAS-Based Multispectral and Hyperspectral Imagery. Remote Sensing **16**(13), 2363 (Jan 2024). https://doi.org/10.3390/rs16132363, number: 13 Publisher: Multidisciplinary Digital Publishing Institute 2
4. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In: Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://papers.nips.cc/paper_files/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html 1
5. Li, J., Chen, P., Yu, S., He, Z., Liu, S., Jia, J.: Rethinking Out-of-distribution (OOD) Detection: Masked Image Modeling is All You Need (Apr 2023). https://doi.org/10.48550/arXiv.2302.02615, arXiv:2302.02615 [cs] 1, 2
6. Müllerová, J., Brundu, G., Große-Stoltenberg, A., Kattenborn, T., Richardson, D.M.: Pattern to process, research to practice: remote sensing of plant invasions. Biological Invasions (Aug 2023). https://doi.org/10.1007/s10530-023-03150-z 1
7. Raine, S., Marchant, R., Kusy, B., Maire, F., Fischer, T.: Image Labels Are All You Need for Coarse Seagrass Segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5943–5952 (2024), https://openaccess.thecvf.com/content/WACV2024/html/Raine_Image_Labels_Are_All_You_Need_for_Coarse_Seagrass_Segmentation_WACV_2024_paper.html 1
8. Soviany, P., Ionescu, R.T., Rota, P., Sebe, N.: Curriculum Learning: A Survey. International Journal of Computer Vision **130**(6), 1526–1565 (Jun 2022). https://doi.org/10.1007/s11263-022-01611-x 1
9. Valdenegro-Toro, M.: Sub-Ensembles for Fast Uncertainty Estimation in Neural Networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4119–4127 (2023), https://openaccess.thecvf.com/content/ICCV2023W/LXCV/html/Valdenegro-Toro_Sub-Ensembles_for_Fast_Uncertainty_Estimation_in_Neural_Networks_ICCVW_2023_paper.html 1, 2
10. Wickramasinghe, C.S., Marino, D.L., Manic, M.: ResNet Autoencoders for Unsupervised Feature Learning From High-Dimensional Data: Deep Models Resistant to Performance Degradation. IEEE Access **9**, 40511–40520 (2021). https://doi.org/10.1109/ACCESS.2021.3064819, conference Name: IEEE Access 1