

Adapting Vision Transformers for High Resolution Hyperspectral Tree Trait Estimation

Julius Pesonen
Finnish Geospatial Research
Institute and Aalto University

Sophie Fabre
ONERA

Eija Honkavaara
Finnish Geospatial Research
Institute

Summary

We propose a method for adapting a pretrained vision transformer (ViT) for estimating bio-chemico-physical parameters of trees. The adaption requires generating synthetic data, which is done using a separate 1D spectrum-to-traits model as shown in **Figure 1**. (on the right). The adapted model acts as an effective baseline for analysing alternative ways of training models for the task.

Hyperspectral AVIRIS data

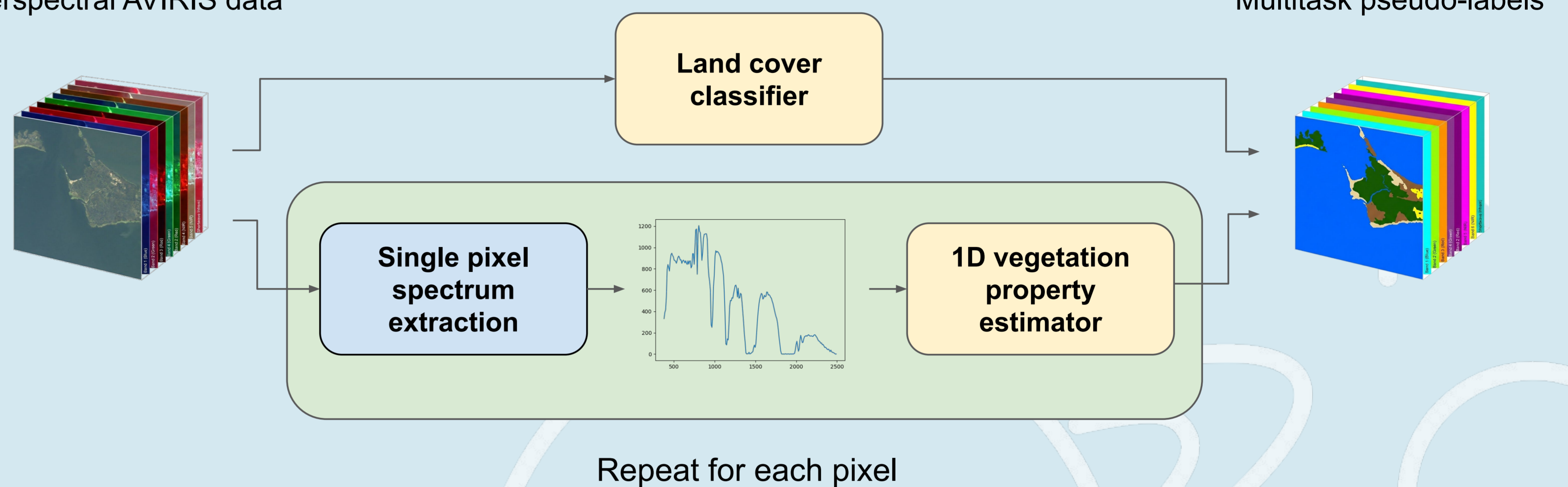


Figure 1. The pseudo-labelling process for generating training data for the model adaption.

Challenge

- Previous methods for estimating bio-chemico-physical parameters of trees act on individual spectra
- Using image-to-image models enables leveraging information from large-scale pretraining via transfer learning
- Adapting the ViT for the task requires synthesising training data as existing data for the task only links individual traits and spectra

Training data

- AVIRIS data from various locations in the US and the Pacific
- Spatial resolution 4–20 metres
- 224 spectral bands in the range of 380–2500 nm
- samples visualised on the left (**Figure 2**.) with the pseudo-labels

Test data

- Test site in Southern France with 35 pine trees analysed for model evaluation
- Lab measurements of water content, chlorophyll concentration (chlorophyll a + b), and carotenoid concentration (β -carotene + lutein + antheraxanthin)
- Imaged with 0.75 metre spatial resolution
- 420 spectral bands in the range of 380–2500 nm
- Visualised below (**Figure 3**.) with test trees highlighted in white

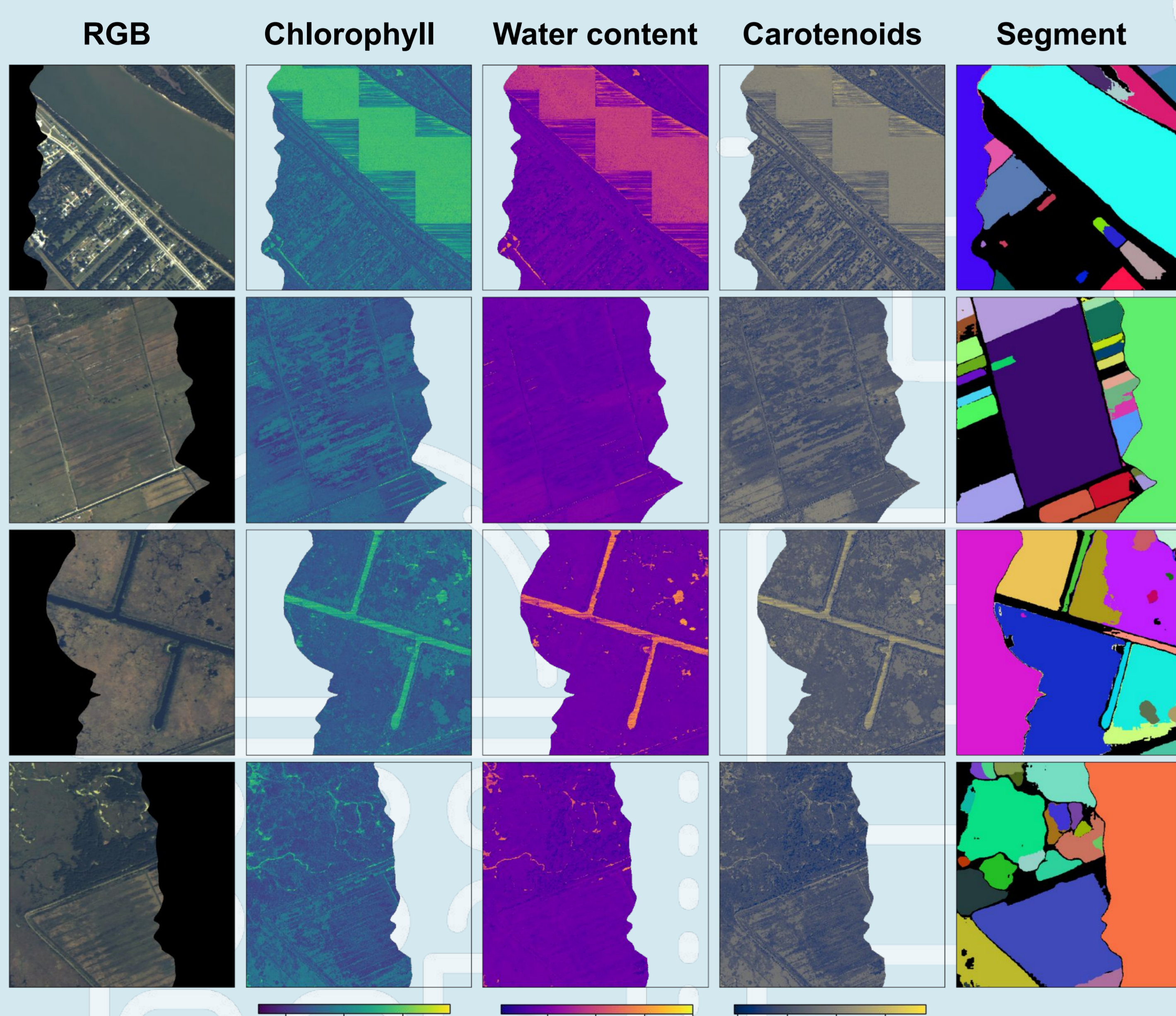


Figure 2. Vegetation parameter pseudo-labels generated using the 1D model. The land cover instance segments are from the pretrained ViT.

Methods

- We leverage an existing pixel-wise model [1] to pseudo-label AVIRIS imagery
- The AVIRIS imagery is used to fine-tune a ViT model pretrained with large scale land cover classification data [2]
- The resulting model combines information from the original ViT pretraining and that of the 1D trait estimation model



Figure 3. The test area visualised in RGB with the test tree segments highlighted in white.

Conclusion

We present an alternative method for obtaining simultaneous land cover classification with biochemical and physical parameter estimates from a single model, using a complete hyperspectral image as an input. Our method promises improvements in predictions by incorporating richer spatial information from multiple pixels at a time as well as leveraging large-scale pretraining for additional robustness.

References

- [1] Cherif, E., Feilhauer, H., Berger, K., Dao, P. D., Ewald, M., Hank, T. B., ... & Kattenborn, T. (2023). From spectra to plant functional traits: Transferable multi-trait models from heterogeneous and sparse data. *Remote Sensing of Environment*, 292, 113580.
[2] Li, J., Liu, Y., Wang, X., Peng, Y., Sun, C., Wang, S., ... & Zhong, Y. (2025). HyperFree: A channel-adaptive and tuning-free foundation model for hyperspectral remote sensing imagery. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 23048-23058).