Application of SATLAS Foundation Model to increase productivity on European Small Woody Features detection

Augot J.[1], Poitevin E.[1], Meda G.[1], Carpentier B.[1], Faucqueur L.[1], Steunou K.[1]

1. Collecte Localisation Satellites (CLS), 11 rue Hermès, 31520 Ramonville-Saint-Agne, France

Introduction

Foundation Models represent a significant advancement in machine learning, offering robust capabilities for a wide range of applications. In this study, we used Satlas Foundation Model for automatic classification of Small Woody Features (SWF), the data product produced under the Copernicus Land Monitoring Service (CLMS), leveraging its advanced architecture and pretrained weights.

To comprehensively evaluate its performances, we benchmarked Satlas against a U-Net on our SWF dataset. More precisely, we trained both models with various number of training samples, to quantify the amount of labelled data necessary to reach the desired metrics.

Foundation Model Architecture

The classification of Small Woody Features from VHR images is based on Satlas Foundation Model (Bastani et al, 2023). This FM is pretrained on a large-scale dataset called *SatlasPretrain* (Bastani et al, 2023), which combines worldwide aerial and satellite images like Sentinel-2 and NAIP, with 7 sources of labels for land classification.

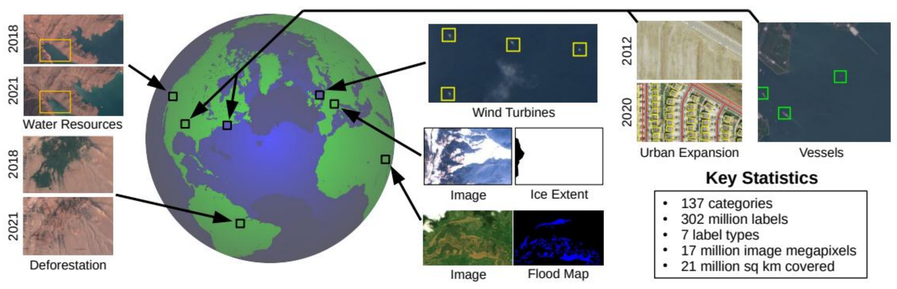

Fig 3: Overview of SatlasPretrain dataset

Satlas model consists of three main components: a backbone, a Feature Pyramid Network (FPN), and a prediction head that classifies images for a specific task.

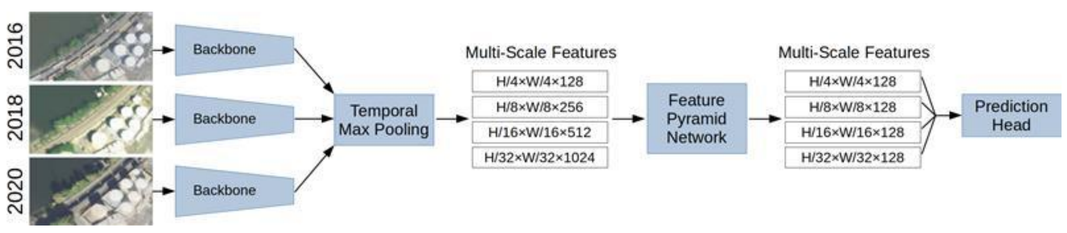

Fig 4: Satlas model architecture

Of the components shown in figure above, the backbone + FPN of the aerial pre-trained model (0.5-2m/px high-res imagery, Single-image, RGB) were used, then the segmentation prediction head was attached and fine-tuned on our SWF database.

For finetuning, SWF input images have 4 bands: Red, Green, Blue and Near Infrared (NIR). Originally, pre-trained Satlas backbones can only handle 3 bands (RGB). To include the NIR in the model without having to retrain the backbone, we concatenated it to the feature maps at the end of the FPN, before passing the whole to the prediction head. This showed an improvement in the classification accuracy of vegetation.

References

Papers:

Bastani, F., Wolters, P., Gupta, R., Ferdinando, J., & Kembhavi, A. (2023). Satlaspretrain: A large-scale dataset for remote sensing image understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 16772-16782).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer international publishing.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Small Woody Features Use Case

Small Woody Features are vital components of the European landscape. They provide numerous ecological benefits, including acting as habitat oases, enhancing biodiversity and helping to fight against soil erosion.

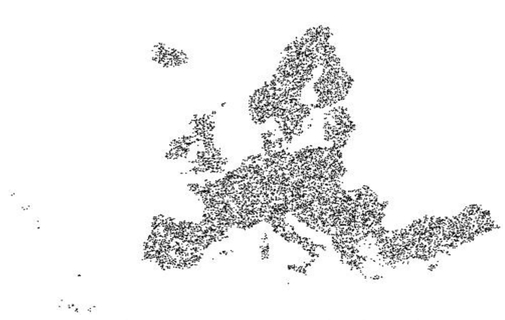

Fig 1: Dataset samples location

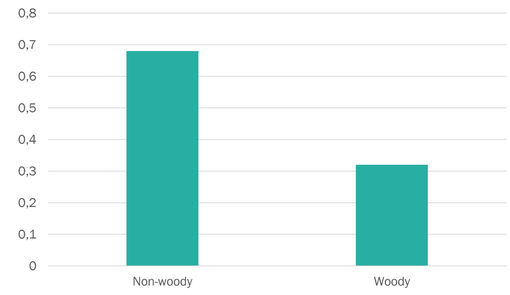

Fig 2: Dataset classes distribution

In the framework of the Copernicus Land Monitoring Service, the High-Resolution Layer Small Woody Features is updated every three years.

Benchmark and Results

U-Net

A U-Net (Ronneberger et al, 2015) with Resnet50 (He et al, 2015) encoder was chosen as baseline comparison in this benchmark.

Influence of the amount of training samples

Satlas and U-Net were trained with 5, 50, 500, 5000, 10000 labelled images, and tested on the same 1000 independent test images.

# training samples	5			50			500			5000			10000		
Model	Satlas	U-Net	Diff	Satlas	U-Net	Diff	Satlas	U-Net	Diff	Satlas	U-Net	Diff	Satlas	U-Net	Diff
Acc	**0.789**	0.645	14.4%	**0.846**	0.822	2.4%	**0.888**	0.860	2.8%	**0.890**	0.867	2.3%	**0.884**	0.878	0.6%
F1	**0.769**	0.632	13.7%	**0.826**	0.800	2.6%	**0.873**	0.841	3.2%	**0.875**	0.852	2.3%	**0.869**	0.864	0.5%
IoU	**0.539**	0.393	14.6%	**0.623**	0.579	4.4%	**0.706**	0.649	5.7%	**0.711**	0.672	3.9%	**0.702**	0.698	0.4%

Table 1: Satlas VS U-Net trained on various number of training samples

Conclusions:

1. Satlas overperforms the U-Net in all cases. The fewer training images, the bigger the gap between them. Satlas starts performing well with only 5 training images, while the U-Net performances increase with the number of samples, equalling Satlas with 10000 training samples.
2. **Satlas performs better when trained with only 500 images than U-Net with 10000 (red boxes).**

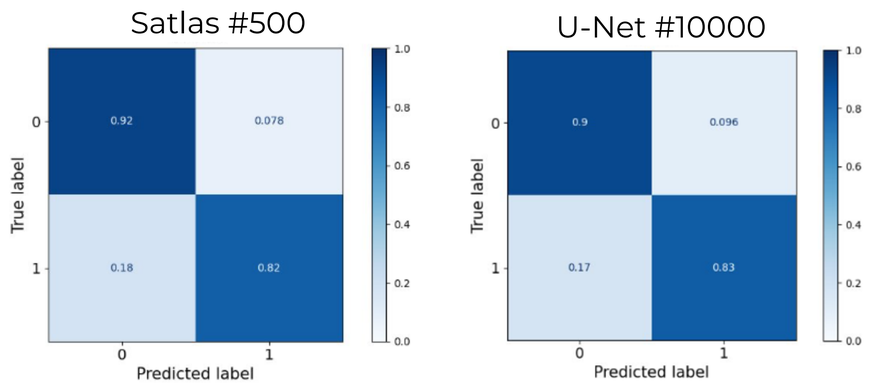

Fig 5: Confusion matrices of Satlas trained on 500 samples, and U-Net on 10000.

Visualisations on test set

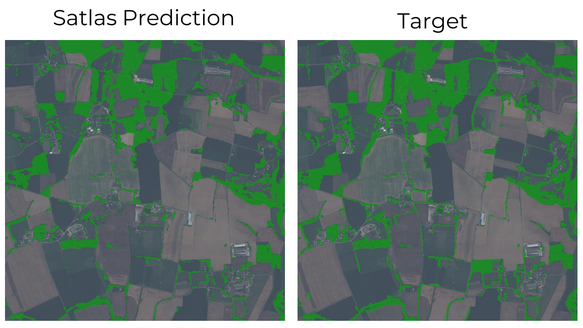

Fig 6: Sample 1 visualisation

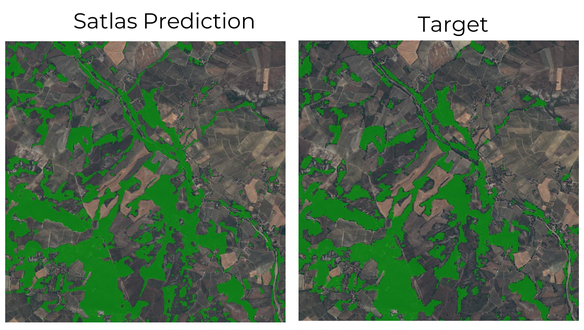

Fig 7: Sample 2 visualisation

Conclusion

We benchmarked Satlas Foundation Model against a U-Net on our Small Woody Features dataset. Satlas trained with only 500 labelled images performs better than the U-Net trained on 10000 samples. **This shows that using a FM like Satlas can reduce by 20 the number of annotations needed to automatically classify woody features.**

This study was done with only 2 classes, but the same benchmark is being done with a more complex dataset (12 classes) to see if similar conclusions will be drawn.