LEVERAGING HISTORICAL AND RADAR SATELLITE DATA FOR LEAF AREA INDEX ESTIMATION WITH DEEP LEARNING

Clement Wang, Antoine Debouchage*, Valentin Golditè*, Aurélien Wery, Jules Salzinger***

*CentraleSupélec - University of Paris-Saclay, **Austrian Institute of Technology - *Vienna, Austria*

ABSTRACT

The Leaf Area Index (LAI) is a critical parameter to understand ecosystem health and vegetation dynamics. In this paper, we propose a novel method for pixel-wise LAI prediction by leveraging the complementary information from Sentinel 1 radar data and Sentinel 2 multi-spectral data at multiple timestamps. Our approach uses a deep neural network based on multiple U-nets tailored specifically to this task. To handle the complexity of the different input modalities, it is comprised of several modules that are pre-trained separately to represent everything in a common latent space. Then, we fine-tune them end-to-end with a common decoder that also takes into account seasonality, which we find to play an important role. Our method achieved 0.06 RMSE and 0.93 R^2 score on publicly available data. We make our contributions available[1] for future works to further improve on our current progress.

***Index Terms*—** Leaf area index, De-clouding, Deep learning, U-net, Pixel-wise regression, Remote Sensing

1 Introduction

The Leaf Area Index (LAI) is a fundamental vegetation parameter that quantifies the total area of leaves per unit ground area. It serves as a key indicator of plant productivity, energy exchange processes, and overall health of the ecosystem [1]. Accurate estimation of the LAI is essential for various applications, including ecological modeling, crop yield prediction, carbon cycle assessment, and climate change studies.

Traditionally, the LAI has been estimated using labor-intensive and time-consuming methods, such as destructive sampling or indirect measurements based on allometric equations [2] [3]. While these approaches provide valuable insights, they are often limited in their spatial coverage and representativeness.

As a result, remote sensing data from satellite platforms have gained significant attention as a valuable source for LAI predictions. Satellite-based observations offer the advantage of providing repetitive coverage over large areas, enabling the assessment of LAI dynamics at regional to global scales. However, despite their many advantages, satellite data face significant challenges, notably the interference caused by clouds in accurate LAI estimation. Cloud cover poses a substantial obstacle to LAI assessment, as it obstructs direct measurements and reduces the quality and availability of cloud-free observations. Overcoming this challenge is critical to leveraging the full potential of satellite data for comprehensive and reliable LAI prediction.

Our approach focuses on pixel-wise LAI prediction using deep learning. Deep learning has demonstrated its ability to extract useful information and learn complex patterns from extensive datasets. To the best of our knowledge, this study represents a pioneering investigation into the prediction of LAI by harnessing the combined information from Sentinel 1 and Sentinel 2 data at multiple timestamps with a deep neural network.

We train our method on public Sentinel-1 and 2 LAI data in different areas in Europe. Importantly, we leverage the full diversity of this data to test it in 3 different contexts: for areas around the training areas and devoid of clouds, for areas around the training areas with cloudy past information, and for new areas far from any training location. We show that the proposed method demonstrates promising results in all settings.

2 RELATED WORKS

Satellite data has gained popularity due to its convenience in LAI prediction, offering significant advantages over direct methods or allometric approaches [2] [3]. Moreover, as demonstrated by [4] and [5], satellite data exhibit comparable performance to traditional alternatives. In certain cases, it can even outperform them in explainability, particularly when supplemented with additional metadata such as terrain variables.

Most studies on LAI prediction from satellite data have primarily focused on specific settings, often limited to a single type of crop and to the optical data from Sentinel 2 or Landstad 8. These studies commonly employ various methods, including multi-regression [5] [6] [7] [8], Fully connected neural networks [8] [9], Bayesian networks [10], as well as classical machine learning techniques [11] [12] such as gradient boosting, Gaussian process, support vector machines (SVM),

[1] https://github.com/valentingol/LeafNothingBehind

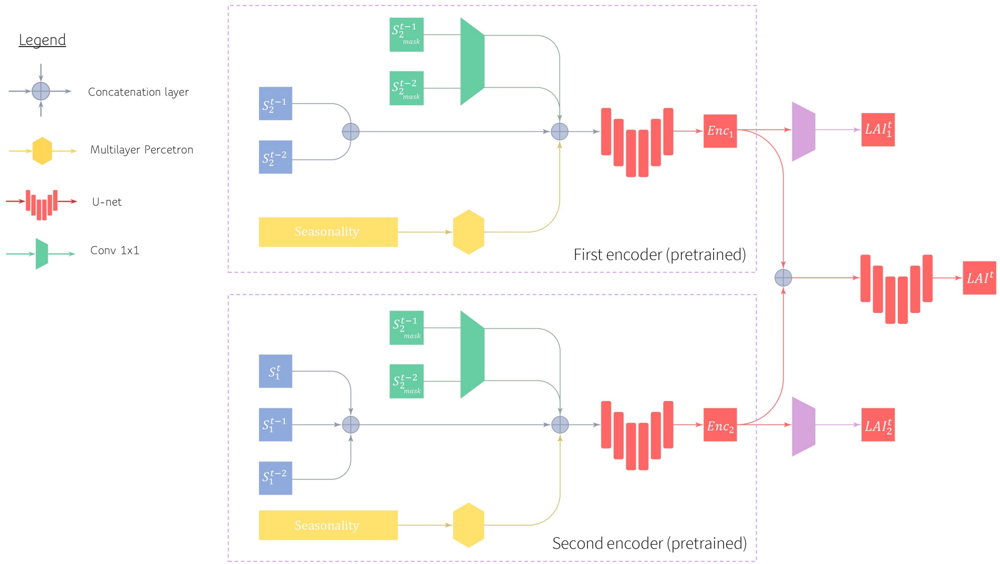

Fig. 1. Our architecture

and random forest. [13] adopts Gaussian processes on Sentinel 2 and Sentinel 1 data at multiple timestamps, resembling our approach. A frequent limitation in prior work is that each pixel is treated independently, overlooking the valuable contextual information from neighboring pixels. This disregard for the global context surrounding each pixel is a missed opportunity that warrants further exploration and consideration.

In parallel, cloud removal has garnered significant attention within the realm of deep learning research. Numerous studies have explored the development of robust methods for cloud removal on satellite data using deep learning techniques. In particular, more powerful convolutional architectures have been investigated, capitalizing on the convolutional inductive bias that takes advantage of neighboring pixel information. Prominent architectures such as U-Net [14] and ResNet [15] have been widely employed in addressing the challenge of cloud removal. These architectures offer sophisticated and effective mechanisms for capturing spatial dependencies and contextual information, enabling more accurate and efficient cloud removal algorithms. Most studies rely on both radar and optical data at one single timestamp or several timestamps [16].

3 METHOD

3.1 Dataset

In this study, we use publicly available LAI data extracted from Sentinel 2 data in different European countries (Bulgaria, Denmark, France, Germany, Lituania, Moldavia, Spain, Sweden, Ukraine, and the United Kingdom). We pair this data with Sentinel 1 radar data captured no more than a day from the passage of Sentinel 2. Each sample corresponds to 3 consecutive Sentinel 2 LAI maps S_2^T, pixel-aligned with corresponding Sentinel 1 radar data images S_1^T, with $T \in \{t-2, t-1, t\}$. Those three timestamps are separated by no more than a week. This data is processed into images of 256×256 pixels. We also have access to semantic masks ${S_2}_{mask}^T$ describing the nature of the element observed by Sentinel 2 (e.g., cloud, water, land, etc.). Finally, the date of each observation is also available and plays a significant role in our method. Our goal is to use these various clues to predict the last LAI image of each time series.

For testing purposes, we use randomly selected 256×256 squares extracted from the aforementioned locations and select a first set of those where past observations are particularly clear, then a second where they are particularly cloudy. As a measure against overfitting at training time, we validate our model on squares located close to the test squares during development. To ensure we do not suffer from validation overfitting, we complement these two test sets with a dataset from different locations, namely in the Czech Republic and Italy. We later refer to those datasets as "non-cloudy", "cloudy" and "unique areas" respectively.[2]

3.2 Architecture

3.2.1 Description

Our proposed architecture consists of three distinct components: two parallel encoders and one decoder (Figure 1).

The first encoder processes the input Sentinel 1 radar data, mask data, and seasonality information. The masks, which are pixel-wise one-hot-encoded maps, are fed to a point-wise convolution to reduce their dimensionality. The seasonality

[2] we are grateful to World from Space for their help in acquiring, filtering and processing the data

Table 1. Ablation studies metrics

Input data	Architecture	Non cloudy		Cloudy		Unique areas	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
S_1	Encoder 1	0.163	0.434	0.279	0.152	0.260	0.125
$S_1 + S_{2masks}$	Encoder 1	0.117	0.709	0.245	0.345	0.208	0.440
$S_1 + S_{2masks} + seas.$	Encoder 1	0.111	0.735	**0.233**	**0.409**	0.198	0.491
$S_2 + S_{2masks}$	Encoder 2	0.090	0.827	0.253	0.301	0.103	0.863
$S_2 + S_{2masks} + seas.$	Encoder 2	0.067	0.903	0.344	-0.294	0.111	0.839
All	Final	**0.058**	**0.930**	0.238	0.383	**0.101**	**0.867**

information, represented by sine and cosine values encoding the period of the year, is fed through a multi-perceptron module before being broadcast into a feature map. This feature map is then concatenated with the other inputs and collectively passed through the U-net encoder. The second encoder is responsible for processing the input of LAI values at timestamps t-2 and t-1, as well as mask data and seasonality information. Similar to the first encoder, it follows an identical architecture, leveraging the same set of operations and modules.

The decoder component of our architecture receives the concatenated output of the two encoders as input. It is designed as a straightforward U-net structure, which facilitates the integration of the analyzed features from both the Sentinel 1 data and the Sentinel 2 data. By merging the analyzed features from both sources, our model achieves a high level of accuracy in LAI prediction, leveraging the complementary information provided by Sentinel 1 and Sentinel 2 data.

3.2.2 Architecture design

The U-net modules employed in our architecture have been widely recognized for their effectiveness in pixel-wise predictions, such as semantic segmentation tasks [14]. These modules offer several advantages, including a relatively small number of parameters, the incorporation of skip connections for enhanced convergence, and the ability to capture both local and global information.

The overall design of our global architecture works as an inductive bias, directing the flow of information through the network. In line with this design, our approach prioritizes the removal of different timestamp data in the initial stages. This decision is motivated by the understanding that retaining multiple timestamp data can lead to an unnecessary increase in the number of feature maps and potential redundancy when cloud cover is absent. By removing redundant timestamp data, we streamline the feature representation and optimize the LAI prediction process.

3.2.3 Intermediate supervision and pre-trained weights

Given the complexity of our architecture, we adopted a two-step training approach. First, we trained each encoder separately by introducing a pixel-wise convolution layer and optimizing the mean squared error (MSE) loss for each encoder. This initial training allowed us to obtain pre-trained weights for each encoder.

Subsequently, we incorporated these pre-trained weights into the entire architecture, utilizing intermediate supervision to optimize the overall model. Intermediate supervision involves optimizing the loss at intermediate stages, ensuring that the model maximizes the utilization of each input modality. This approach not only enables the model to make the most of the pre-trained weights' feature representation but also helps preserve and enhance the learned features during the early stages of training. The loss between the ground truth LAI_t^{gt} and $pred = \{LAI_t^{dec}, LAI_t^{enc_1}, LAI_t^{enc_2}\}$ is expressed as:

$$\begin{aligned} loss(LAI_t^{gt}, pred) &= MSE(LAI_t^{gt}, LAI_t^{dec}) \\ &+ \alpha.MSE(LAI_t^{gt}, LAI_t^{enc_1}) \\ &+ \beta.MSE(LAI_t^{gt}, LAI_t^{enc_2}) \end{aligned}$$

α and β are hyper-parameters to weigh down the importance of intermediate stages compared to the last stage. Note that in practice, the loss only takes into account pixels that are not covered by clouds, which we extract from S_{2mask}^t.

4 EXPERIMENTS

4.1 Implementation details

The implementation was made in Pytorch. We trained with Adam for 100 epochs with an initial learning rate of 0.001. To enhance training stability and performance, we incorporated learning rate decay, applying two equally spaced decays with a decay factor of 0.2. For the loss parameters in the intermediate supervision, we used $\alpha = 0.1$ and $\beta = 0.15$. The batch size was set to 32 and the training was carried out on a NVIDIA GeForce RTX 3090 GPU.

4.2 Results

In order to evaluate the effectiveness of our proposed method and gain deeper insights into the relationship between the input data and various components of the architecture, we conducted a series of ablation studies. These studies involved selectively training specific parts of the architecture and focusing on specific subsets of the input data. When training only on S_1 (resp. S_2), we drop the second (resp. first) encoder and the decoder, and we take the output of the intermediate stage for evaluation. See Table 1.

The high correlation observed between non-cloudy and unique area metrics provides reassurance regarding the performance of our model, both during training and at inference time on unseen locations. However, it is worth noting that the error rates were found to be twice as high in unique areas, indicating that the models excel primarily in trained areas. This suggests that the models might have learned biases specific to the training locations. Therefore, incorporating additional meta-information about the location, such as terrain variables or meteorological data, could potentially enhance the model's performance further. The positive impact of seasonality on the metrics further supports this notion, indicating that the inclusion of additional metadata has the potential to yield even better results.

The evaluation of our models on cloudy data revealed more mixed results. Specifically, we observed that S_1 provided more relevant information compared to S_2. Surprisingly, even our final model incorporating S_1 data exhibited poorer performance on cloudy data. We attribute this discrepancy to the dataset's inherent imbalance concerning cloudy data. The scarcity of cloudy data in the training set may have hindered the models' ability to optimise their reliance on S_1 data when faced with cloudy conditions. To address this challenge, an interesting approach would involve assigning a higher weight to the loss function for cloudy inputs during training.

5 CONCLUSION

Our study has demonstrated the effectiveness of deep learning techniques in predicting LAI by integrating global and local information from both radar and multispectral data sources at several timestamps. Moving forward, there is potential for further improvement by specifically training the model on cloudy data, emphasising the importance of handling data with varying cloud cover.

We also express our gratitude to World from Space for providing the data for this project and for their support.

REFERENCES

[1] Knyazikhin, Y., Schull, M. A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., ... & Myneni, R. B. (2013). Hyperspectral remote sensing of foliar nitrogen content. Proceedings of the National Academy of Sciences, 110(3), E185-E192.

[2] Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & Baret, F. (2004). Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. Agricultural and forest meteorology, 121(1-2), 19-35.

[3] Fang, H., Baret, F., Plummer, S., & Schaepman-Strub, G. (2019). An overview of global leaf area index (LAI): Methods, products, validation, and applications. Reviews of Geophysics, 57(3), 739-799.

[4] Friedl, M. A., Schimel, D. S., Michaelsen, J., Davis, F. W., & Walker, H. (1994). Estimating grassland biomass and leaf area index using ground and satellite data. International Journal of Remote Sensing, 15(7), 1401-1420.

[5] Green, E. P., Mumby, P. J., Edwards, A. J., Clark, C. D., & Ellis, A. C. (1997). Estimating leaf area index of mangroves from satellite data. Aquatic botany, 58(1), 11-19.

[6] Colombo, R., Bellingeri, D., Fasolini, D., & Marino, C. M. (2003). Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote sensing of environment, 86(1), 120-131.

[7] Aboelghar, M., Arafat, S., Saleh, A., Naeem, S., Shirbeny, M., & Belal, A. (2010). Retrieving leaf area index from SPOT4 satellite data. The Egyptian Journal of Remote Sensing and Space Science, 13(2), 121-127.

[8] Richter, K., Hank, T. B., Vuolo, F., Mauser, W., & D'Urso, G. (2012). Optimal exploitation of the Sentinel-2 spectral capabilities for crop leaf area index mapping. Remote Sensing, 4(3), 561-582.

[9] Jin, X., Li, Z., Feng, H., Ren, Z., & Li, S. (2020). Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. The Crop Journal, 8(1), 87-97.

[10] Kalácska, M., Sánchez-Azofeifa, G. A., Caelli, T., Rivard, B., & Boerlage, B. (2005). Estimating leaf area index from satellite imagery using Bayesian networks. IEEE Transactions on Geoscience and Remote Sensing, 43(8), 1866-1873.

[11] Mao, H., Meng, J., Ji, F., Zhang, Q., & Fang, H. (2019). Comparison of machine learning regression algorithms for cotton leaf area index retrieval using Sentinel-2 spectral bands. Applied Sciences, 9(7), 1459.

[12] Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R. B., & Chang, Q. (2019). Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. ISPRS Journal of Photogrammetry and Remote Sensing, 154, 189-201.

[13] Campos-Taberner, M., García-Haro, F. J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Busetto, L., ... & Boschetti, M. (2017). Exploitation of SAR and optical sentinel data to detect rice crop and estimate seasonal dynamics of leaf area index. Remote Sensing, 9(3), 248.

[14] 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.

[15] 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).

[16] Ebel, P., Garnot, V. S. F., Schmitt, M., Wegner, J. D., & Zhu, X. X. (2023). UnCRtainTS: Uncertainty Quantification for Cloud Removal in Optical Satellite Time Series. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2085-2095).