

PlantFusion-1D2D: Leaf Area Index Estimation of Multimodal Deep Learning Framework under Complex Weather Conditions

Shulin Sun, Man Zhang, Hanxu Wang, Jiaren Zhou, Han Li, Ruicheng Qiu, Minjuan Wang*

Key Lab of Smart Agriculture Systems, Ministry of Education,
College of Information and Electrical Engineering, China Agricultural University, Beijing, China
{shulinsun, cauzm, wanghanxu, zhoujiaren, caulihan, qrc, minjuan}@cau.edu.cn

1. Introduction

The leaf area index(LAI) is a key parameter for assessing crop vigor and light-use efficiency. LAI is defined as the total area of all green leaves per unit ground surface area. Accurate LAI detection supports precision agriculture and enhances ecosystem understanding. Traditional LAI measurement methods include direct and indirect methods. Direct measurement is labor-intensive and destructive, while indirect measurement, such as the LAI-2200c, estimates LAI by recording light intensity and calculating canopy interception rates. However, this method can be affected by cloud movement and is costly.

Recent research has utilized remote-sensing techniques for high-resolution LAI retrieval. These methods fall into satellite imagery and UAV-based categories. unmanned aerial vehicles(UAVs), equipped with high-resolution sensors, offer more accurate data but face cost and technical limitations. UAV remote sensing has advanced rapidly but still requires high costs, steep learning curves, and air-space restrictions. Cloud contamination poses a significant challenge to both satellite and UAV remote sensing methods. Thin-clouds alter vegetation spectral reflectance, complicating accurate data acquisition. Recent deep learning advancements have improved decision-making for LAI inversion. Handheld devices, offering portability and low cost, have become popular alternatives to traditional methods.

Exploiting the device's short acquisition time—which minimizes the impact of changing solar illumination—and integrating the thin-cloud spectral image restoration technique proposed here, together with noise-reduction modules, this study applies the PlantFusion-1D2D multimodal network to the handheld-acquired spectral data, RGB data, and vegetation indices selected from characteristic bands. The approach effectively improves the accuracy of LAI inversion for the target plots.

2. Materials and methods

The handheld device used in this study is a cost-effective solution to LAI inversion, offering fast and accurate results. The method integrates spectral data quality enhancement, feature selection, and a thin-cloud-corrected multimodal network for LAI inversion. The aim is to provide a portable solution capable of operating under thin-cloud interference and accurately predicting LAI in rapidly changing field conditions.

Data were collected from rice plots at various growth stages in Hainan, China, using a handheld device. Both spectral and RGB images were collected, with true LAI values measured by the LAI-2200C at plot corners. A total of 414 samples were used, including 67 contaminated by thin-clouds. Data preprocessing involved normalization, outlier detection, and noise reduction to ensure high-quality data for LAI inversion. Statistical methods were employed to identify and correct anomalous data, ensuring stable and accurate inversion.

2.1. Selection of key feature bands

Feature band selection reduced computational burden by optimizing spectral band combinations. A convolutional neural network(CNN) captured nonlinear relationships between spectral data, vegetation indices, and LAI. The selected bands significantly improved computational efficiency and LAI prediction accuracy.

2.2. Repair of cloud occlusion data

During the data acquisition process, the sky intermittently becomes overcast by thin-cloud due to the relatively high thin-cloud content in coastal regions. As these thin-cloud do not cast pronounced shadows on the ground, researchers are unable to detect their presence in real time. It is often only during data processing that significant issues with the spectral energy values of the affected plots are discovered. To avoid wasting data compromised by thin-cloud,

we propose a fully connected neural network based on the Multi-Head Self-Attention (MHSA) mechanism for correcting such data. By training on both clear-sky data and anomalous thin-cloud data, we utilize a dataset comprising 67 samples contaminated by thin-cloud and 270 sets of normal data captured at similar times without thin-cloud contamination. This dataset is partitioned into 80% for training and 20% for validation. Taking 27 spectral bands as input, we project them through a linear layer into a hidden layer containing 64 neurons and employ the Rectified Linear Unit (ReLU) activation function. Simultaneously, a multi-head attention mechanism is introduced to focus on the complex relationships among bands, thereby enhancing the model's attention to bands polluted by thin-cloud. This enables the model to learn the error range induced by thin-cloud occlusion and the inter-band dependencies within the spectrum, improving the restoration of data affected by thin-cloud.

2.3. PlantFusion-1D2D multimodal network inversion of LAI

We propose a thin-cloud-corrected, multimodal LAI inversion model, termed the PlantFusion-1D2D network, which integrates one-dimensional (1D) and two-dimensional (2D) data. The network comprises two parallel processing branches—spectral and RGB—with their outputs merged via a self-attention module. In the spectral branch, two 1D convolutional layers (each with 64 filters and ReLU activation) process an input vector of 55 dimensions (27 spectral-band reflectance values and 28 vegetation indices). In the RGB branch, resized RGB images (224×224 pixels) are fed into a ResNet-101 model with its classification head removed, yielding a 2048-dimensional feature vector. To fuse these heterogeneous features, we concatenate the 128-dimensional spectral output with the 2048-dimensional RGB output and apply a four-head multi-head self-attention mechanism. By weighting features dynamically, the self-attention module produces a 2176-dimensional fused representation. A single mapping layer then relates this fused vector to the ground-truth LAI, enabling accurate prediction.

3. Results and discussion

3.1. Evaluation of cloud impact data repair results

We conducted experimental comparisons between the corrected and uncorrected data using several machine learning models, with results summarized in the table. Most models show a significant increase in R^2 after correction—for example, the Ridge Regression model's R^2 rose from 0.17 to 0.56, and the SVR (RBF kernel) model's R^2 increased from 0.23 to 0.56. This indicates that the corrected data can be effectively fitted by these models, thereby demonstrating the effectiveness of spectral correction. In the future, expand-

ing the sample size and incorporating temporal dynamics into the modeling process may further enhance model performance and broaden its applicability.

3.2. Evaluation of PlantFusion-1D2D multimodal network

We used a variety of machine learning and deep learning methods, using a multimodal network based on DenseNet, EfficientNet, and VGG16 backbones to fuse RGB imagery with spectral reflectance and vegetation indices. The fusion strategy is the same as PlantFusion-1D2D (spectral branch + image branch + fusion layer).

CatBoost achieved the highest $R^2 = 0.69$, closely followed by K-Nearest Neighbors $R^2 = 0.68$, both outperforming several tree-based methods such as Random Forest, Gradient Boosting, LightGBM and XGBoost. By contrast, PlantFusion-1D2D markedly enhanced prediction accuracy through multimodal fusion, elevating R^2 from 0.69 to 0.81 and obtaining the best results across all four evaluation metrics ($R^2 = 0.81$, RMSE = 0.34, MAE = 0.28, MAPE = 7.94 %), thereby surpassing both the traditional machine learning approaches and the DenseNet-, EfficientNet- and VGG16-based multimodal networks.

This experiment confirms that combining spectral reflectance and vegetation indices with RGB imagery can effectively improve LAI prediction accuracy, validating the efficacy of deep learning-based multimodal networks. Future work should investigate the model's robustness and generalization capacity to accommodate a variety of crop types.

4. Conclusion

The study was validated through the collection of 414 sample sets from rice paddy fields in Hainan (including 67 sets of data contaminated by thin-cloud). To address thin-cloud interference, the research developed a neural network based on a multi-head self-attention mechanism to restore the contaminated spectral data.

A multimodal network, PlantFusion-1D2D, for thin cloud correction in LAI retrieval is proposed. This network extracts features from spectral and RGB branches and fuses them using a self-attention mechanism that dynamically adjusts feature weights. Experimental results show that the model achieves an R^2 of 0.81, a mean absolute error (MAE) of 0.28, and a mean absolute percentage error (MAPE) of 7.94%.

These results confirm the effectiveness of the network in recovering cloud-affected data and improving forecast accuracy. The selected spectral bands and PlantFusion-1D2D network reduce computational costs while improving LAI forecast accuracy. This research provides an innovative solution for precision agriculture, filling a gap in existing technologies.