Integrating Phenological Priors with Deep Spatio-Temporal Features for tree species mapping

Zhanyu Ma[1], Ningning Zhu[1], Zhen Dong[1], Ruibo Chen[2], Bisheng Yang[1,*], Zhe Chen[1], Chen Long[1], Ruifei Ding[1]

[1] LIESMARS, Wuhan University, Wuhan, China, 430079 - (zyma2000, ningningzhu, dongzhenwhu, bshyang, ChenZhe_WHU, chenlong107, dingrf)@whu.edu.cn,
[2] Guangxi Zhuang Autonomous Region Institute of Natural Resources Remote Sensing, Nanning, China, 530023 - ruiboil@163.com

Keywords: Tree species mapping, SITS, Prior knowledge, Phenology, Vision Transformer.

Abstract

Mapping large-scale tree species distributions is essential for accurately estimating forest carbon storage. Previous studies have shown that Satellite Image Time Series (SITS) can be effective for classifying tree species. However, many of these studies rely heavily on manual feature engineering or overlook critical geoscientific and forestry knowledge. Such domain-specific insights are particularly important in Earth observation because the same species can exhibit diverse spatio-temporal behaviors across different regions, leading to lower accuracy and limited model robustness. In this work, we propose a novel model, PTSViT, which integrates phenological information with deep spatio-temporal features to address these limitations. Our model's loss function incorporates phenological priors, utilizing ground-based phenological data and tree species labels as supervisory signals to guide the learning of spatio-temporal encoders. We evaluate PTSViT on a newly created dataset, GXData, which includes 11 major tree species in GuangXi. Our model surpasses previous approaches across all evaluation metrics, demonstrating the value of integrating prior knowledge for automated, accurate tree species mapping.

1. Introduction

Tree species mapping is essential for accurately assessing forest biomass and carbon storage, both of which are critical indicators of the global response to climate change. Detailed information on tree species distribution not only enhances insights into forest health and biodiversity but also forms the basis for precise calculations of carbon storage within forest ecosystems(Houghton, 2005). Traditional methods for mapping tree species distribution have primarily relied on field surveys and manual interpretation. While these approaches can produce accurate results, their high time and labor costs limit their applicability over large areas. Additionally, the coverage and accuracy of field surveys are often constrained by the researchers' expertise and workload, making comprehensive and efficient forest monitoring a considerable challenge(Foody, 2002).

With advances in remote sensing technology, the availability of extensive remote sensing data has significantly reduced the cost and complexity of mapping tree species. Data from different modalities, such as satellite imagery, hyperspectral imagery, and LiDAR, provides rich surface information, enabling scientists to efficiently classify tree species(Fassnacht et al., 2016). Current tree species mapping approaches typically operate at two scales: small scale (e.g., forest stands) and large scale (e.g., provincial or national levels). Small-scale tree species classification often utilizes hyperspectral imagery, high-resolution imagery, or LiDAR data, which can deliver highly detailed surface information and, consequently, higher classification accuracy(Fassnacht et al., 2016). However, the processing costs of these data types are considerable, limiting their scalability for large areas(Lu and Weng, 2007). As a result, while small-scale mapping approaches achieve high classification accuracy, they are often confined to localized studies and pose challenges for large-scale forest biomass assessments.

In contrast, large-scale tree species classification typically utilizes medium-resolution satellite imagery, such as Landsat and Sentinel. These images cover extensive areas, enabling tree species mapping at provincial or national scales and providing valuable data for forest biomass estimation(Wulder et al., 2016). However, the relatively low spatial resolution of medium-resolution imagery restricts the ability to accurately distinguish between tree species, resulting in decreased classification accuracy. Additionally, classification robustness is challenged when dealing with complex terrain and diverse forest ecosystems(Kanan et al., 2023). Consequently, enhancing the accuracy and robustness of large-scale tree spe-cies classification remains a critical focus of ongoing research.

For large-scale tree species mapping, two primary approaches are utilized: traditional machine learning methods and deep learning methods. Traditional machine learning methods, such as Random Forest and Support Vector Machine, typically integrate topographic, phenological, environmental, and multi-temporal data to enhance classification accuracy(Grabska et al., 2020)(Kollert et al., 2021)(Balestra et al., 2021). However, these methods rely on feature engineering, which can be time-consuming and labor-intensive. With the expanding volume of remote sensing data, manual feature extraction becomes increasingly impractical.

In contrast, deep learning methods have gained great success due to their ability to automatically extract features. These methods learn complex spatiotemporal features from multi-temporal imagery, avoiding the complicated processes of feature construction and selection. For instance, (Tarasiou et al., 2023) introduced a Vision Transformer architecture specifically designed for satellite image time series, effectively capturing species-specific spatiotemporal features through self-attention mechanisms, thereby reducing reliance on manual feature engineering.

Nevertheless, deep learning methods often overlook domain-specific knowledge, which can lead to less interpretable processes and affect model performance. In particular, in Earth observation domain, the dynamic nature of the environment means that tree species distributions may be influenced by vari-

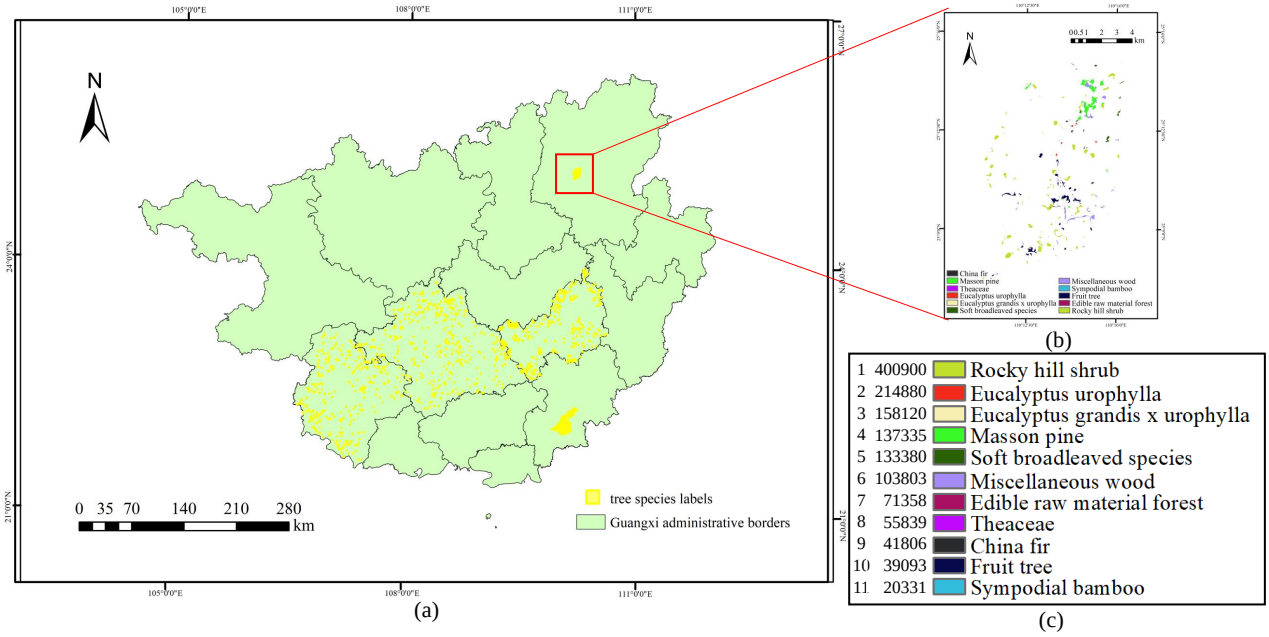

Figure 1. Study area and data description. (a)Location of GuangXi and distribution of tree species labels in GXData; (b)Tree species labels are pixel-level, which are made of a number of polygons; (c) GXData includes 11 tree species and the number of different tree species labels(pixel-level) are shown.

ous factors. Environmental conditions vary by region, resulting in distinct spectral-temporal behaviors for the same species across different areas. By integrating prior knowledge into data-driven models, such as incorporating intrinsic phenological patterns of tree species, classification accuracy and robustness could be significantly improved.

To address the lack of prior knowledge in deep learning-based tree species classification models, we propose PTS-ViT that integrates phenological information with deep spatio-temporal features, overcoming the limitations of interpretability in existing methods and enhancing model performance and robustness. The main contributions of this study can be summarized as follows:

1. We designed a phenology-temporal features alignment module that transforms the output of the temporal encoder and phenological information into a unified one-dimensional time series, providing a consistent representation for heterogeneous data.

2. We developed a phenology-aware loss function, using ground-level phenological information from different tree species as a additional supervisory signal to guide the learning of the temporal encoder, enhancing the model's interpretability.

3. We created a tree species classification dataset based on Sentinel-2, named GXData, which includes 11 dominant tree species within the study area (Guangxi, China). Our method outperforms other comparable methods for the GXData.

2. Related work

2.1 Machine Learning Methods for Tree Species Classification

High temporal resolution satellite imagery, particularly Sentinel-2, have significantly expanded the application of machine learning methods that integrate satellite image time series(SITS) with prior knowledge for large-scale tree species classification. (Balestra et al., 2021) utilized Sentinel-2 satellite imagery from 2018 to 2020 to calculate vegetation indices such as NDVI, TDVI, EVI and GNDVI, and generated time series for forest categories. Machine learning algorithms were trained with accurate ground truth, and PCA was conducted to reduce variable redundancy in the Random Forest classification. Due to the phenological differences among species, the overall classification accuracy was 70% - 80%. (Grabska et al., 2020) evaluated several machine learning algorithms—including Random Forest and Support Vector Machine—using Sentinel-2 imagery combined with environmental data to classify tree species in Polish. Their findings indicate that integrating Sentinel-2 spectral data with environmental variables, such as topography and climate, substantially enhances classification accuracy across various forest species. Similarly, (Kollert et al., 2021) investigated the potential of Sentinel-2 imagery for tree species classification in mountainous regions by leveraging seasonally cloud-free composite images to capture phenological differences. (Hemmerling et al., 2021) utilized all available Sentinel-2 observations and applied radial basis convolutional filters to construct cloud-free 5-day temporal sequences for each spectral band. By inputting these constructed time series into Random Forest classifier, they demonstrated that dense temporal information is critical for improving classification accuracy. In a large-scale study of Canada's forest ecosystems, (Hermosilla et al., 2022) combined Sentinel-2 imagery with environmental data and employed the Random Forest algorithm to predict species distribution. Their results showed that integrating satellite data with environmental variables—such as soil type and climate—can significantly enhance model performance.

Collectively, these studies demonstrate the effectiveness of traditional machine learning algorithms, particularly Random Forest, in leveraging Sentinel-2's high temporal resolutions for large-scale tree species classification. By incorporating auxiliary environmental variables into spatio-temporal features constructed through feature engineering, these models

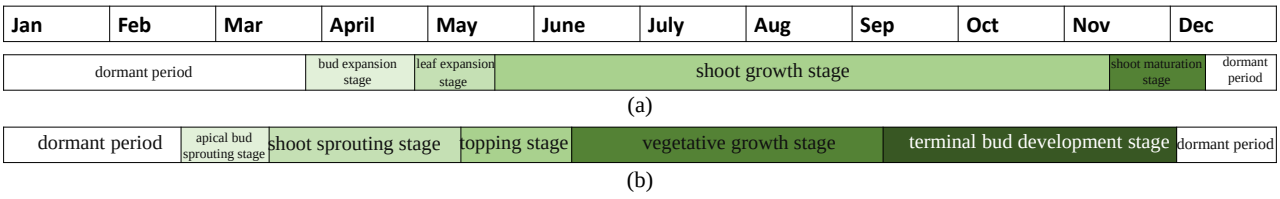

Figure 2. Phenological Periods of Different Tree Species.(a)China Fir; (b)Masson Pine.

achieve higher classification accuracy.

2.2 Deep Learning Methods for Satellite Image Time Series Semantic Segmentation

In recent years, researchers have found that deep learning methods designed for satellite image time series (SITS) can more effectively perform various semantic segmentation tasks, such as land cover classification and crop identification. (Rußwurm and Körner, 2018) applied the ConvLSTM(Shi et al., 2015) model based on Recurrent Neural Networks for land cover classification, utilizing a temporal encoder to capture temporal features within time-series data. This work demonstrates the advantages of RNN in handling SITS. (Martinez et al., 2021) proposed the Fully Convolutional Recurrent Network (FCRN), which combines Convolutional Neural Networks with RNN to process SITS in crop identification tasks. This model extracts spatial features through CNN and temporal features through RNN, significantly enhancing semantic segmentation performance on SITS.

More recently, attention mechanisms have achieved great progress in deep learning. (Garnot and Landrieu, 2021) introduced lightweight temporal self-attention mechanisms(Garnot and Landrieu, 2020) to improve model comprehension of SITS. Their work focused on leveraging self-attention to capture key features in time-series data. (Tarasiou et al., 2023) proposed the TSViT model, a Vision Transformer-based architecture for processing SITS. By using self-attention, the model captures complex spatio-temporal dependencies and demonstrates outstanding performance in semantic segmentation tasks. Transformer has become a promising architecture in processing SITS with their powerful spatio-temporal feature extraction and sequence data modeling capabilities.

3. Materials

3.1 Study Area

Guangxi, located in southwestern China, is characterized by abundant natural resources and diverse ecosystems, covering an area of approximately 236,700 square kilometers (Fig.1(a)). Its topography is predominantly mountainous and hilly. Recent statistics indicate that forest coverage rate of GuangXi has been steadily increasing, with forests now comprising about 60 percents of the region. This increase not only reflects commitment to ecological conservation of the government but also provides essential data support for ecosystem research within the region.

The primary tree species within forest ecosystems of GuangXi include pine, Chinese fir, eucalyptus, and oil tea trees. These species are valuable not only economically but also for their roles in maintaining the stability and health of the ecosystem. Pine and Chinese fir are the main commercial timber species in Guangxi, widely used in construction, paper production, and furniture manufacturing. Oil tea trees provide high-quality edible oil for local communities and contribute positively to environmental protection, promoting biodiversity conservation within the region.

3.2 Data

To minimize the need for repeated atmospheric correction, we directly utilized the Sentinel-2 Level 2A surface reflectance (SR) products obtained from the official European Space Agency website. The Sentinel-2 mission comprises two satellites—Sentinel-2A and Sentinel-2B—launched in 2015 and 2017, respectively. Each satellite has a 10-day revisit interval, resulting in an effective observation frequency of every 5 days over the study area. Sentinel-2 provides 13 spectral bands covering wavelengths from the visible to the shortwave infrared, with spatial resolutions ranging from 10 to 60 meters. For our analysis, we collected imagery data spanning from January 1, 2019, to December 31, 2019, encompassing the various growing seasons of different tree species. To mitigate the impact of clouds and cloud shadows, we selected images with a cloud cover of less than 50 percents. Specifically, we utilized ten spectral bands in this study: blue (B), green (G), red (R), red-edge 1 (R1), red-edge 2 (R2), red-edge 3 (R3), near-infrared 1 (NIR), near-infrared 2 (NIRn2), shortwave infrared 1 (SW1), and shortwave infrared 2 (SW2), while excluding the 60-meter resolution bands.

The species label data used in this study is derived from forest resource surveys (Fig.1(b)), which are periodically conducted by government agencies to assess the status of forest ecosystems. During these surveys, field teams collect data on tree species distribution within the study area. Utilizing publicly available multispectral satellite imagery and species label data, we developed a tree species classification dataset, GXData, which comprises 11 dominant tree species in Guangxi. The overview of the dataset is shown in Fig.1.

This study also incorporates phenological information for different tree species. These phenological data are generally obtained from ground observation stations(Du et al., 2019)(Qin, 1997). Currently, phenological information for Chinese fir and masson pine in Guangxi is directly accessible, as shown in Fig.2.

4. Methodology

We proposed a tree species classification model, PTSViT (Fig.3), which Integrates phenological information with deep spatio-temporal fusion features. PTSViT mainly comprises the following modules: temporal encoder, spatial encoder-decoder, phenology-temporal feature alignment module, and a loss function incorporating phenological priors. The temporal encoder utilizes a self-attention mechanism to construct temporal features from SITS. The spatial encoder encodes the spatial relationships of the images using the fused temporal features from temporal encoder and utilizes a decoder for semantic segmentation tasks. These two modules are derived from the TSViT architecture, which has demonstrated strong performance in crop type mapping. In this study, we adapt them for classifying tree species. The phenology-temporal features alignment module converts the outputs of the temporal encoder and phenological information into one-dimensional time series, creating a unified

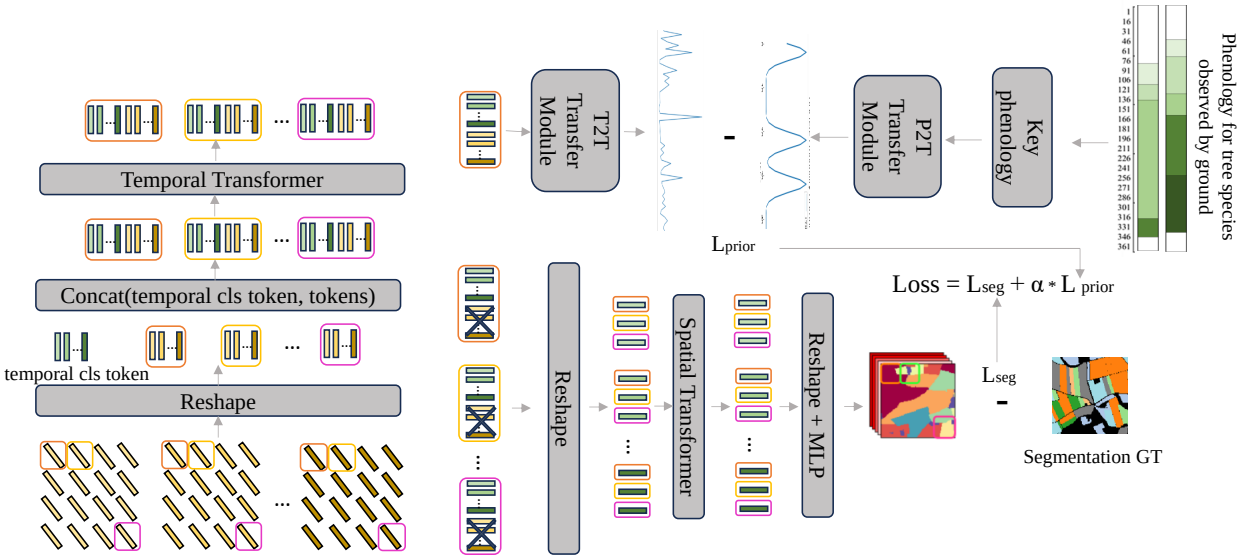

Figure 3. The overall architecture of our model PTSViT. The left part represents temporal encoder, the top right part represents phenology-temporal features alignment module, which includes $T2T$ transfer module and $P2T$ transfer module. The bottom right part represents spatial encoder-decoder.

representation for heterogeneous data. The loss function incorporating phenological priors uses ground phenological information and tree species labels as supervisory signals to guide the learning of the spatio-temporal encoders. The following sections will introduce each module in detail.

4.1 Temporal Encoder

Before processing by the temporal encoder, we first tokenize the SITS record $\mathbf{X} \in \mathbb{R}^{T\times H\times W\times C}$, which consists of a sequence of T temporal satellite images. To adapt the tokenization-as-convolution approach for these records, we apply a 3D kernel of size $(t \times h \times w)$ with strides (t, h, w) across both temporal and spatial dimensions. This process extracts $N = \lfloor \frac{T}{t} \rfloor \lfloor \frac{H}{h} \rfloor \lfloor \frac{W}{w} \rfloor$ non-overlapping tokens $\mathbf{x}_i \in \mathbb{R}^{thwC}$, which are then projected into a d-dimensional space. In our implementation, we set $t = 1$, as in TSViT, so that each token contains only spatial information for each time step.Ultimately, this tokenization scheme is akin to that of ViT, applied in parallel for each acquisition.

After this tokenization scheme, $\mathbf{X}$ are transformed into several tokens, whose size are $(N_T \times N_H \times N_W \times d)$. We reshape it to $\mathbf{Z}_T \in \mathbb{R}^{N_H N_W \times N_T \times d}$ to input to the temporal encoder, as shown in equation(1):

$$\mathbf{Z}_T^0 = \text{concat}(\mathbf{Z}_{T_{cls}}, \mathbf{Z}_T + \mathbf{P}_T[t,:]) \in \mathbb{R}^{N_H N_W \times K+N_T \times d} \quad (1)$$

where $\mathbf{P_T}[\mathbf{t},:] \in \mathbb{R}^{N_T \times d}$ is temporal position encoding in which $\mathbf{t} \in \mathbb{R}^T$ is a vector containing all T acquisition times and $\mathbf{Z_{Tcls}} \in \mathbb{R}^{K\times d}$ is temporal class token, both of which are respectively added and prepended to all $N_H N_W$ time series.

The temporal position encodings $\mathbf{P_T}[\mathbf{t}, :]$ depend dir-ectly on absolute timestep t which introduce acquisition-timespecific biases into the model. For tree species recognition, it is important to encode the absolute temporal position, because it helps model identifying a plant's growth stage.

Consequently, the final feature map of the temporal encoder becomes $\mathbf{Z_T^L} \in \mathbb{R}^{N_H N_W \times K+N_T \times d}$, where the first K tokens in the temporal dimension correspond to the prepended class tokens. We utilize only these first K tokens in the spatial encoder and employ all tokens during the alignment of phenological and temporal features.

4.2 Spatial Encoder-Decoder

Before processing the temporal encoder output by spatial encoder, we exchange the first two dimensions. In this way, we get a list of patch features $\mathbf{Z_S} \in \mathbb{R}^{K\times N_H N_W \times d}$ for all output classes. So the input to the spatial encoder is:

$$\mathbf{Z_S^0} = \text{concat}\,(\mathbf{Z_{Scls}}, \mathbf{Z_S} + \mathbf{P_S}) \in \mathbb{R}^{K\times 1+N_H N_W \times d} \quad (2)$$

Where $\mathbf{P_S} \in \mathbb{R}^{N_H N_W \times d}$ are spatial position encodings and respectively added to all K spatial representations. The output of the spatial encoder is $\mathbf{Z_S^L} \in \mathbb{R}^{K\times 1+N_H N_W \times d}$.

Spatial position encodings $\mathbf{P_S}$ are similar to the position encodings used in the original ViT architecture (Dosovitskiy, 2020), with the difference from temporal position encodings that these biases are now added to K feature maps instead of a single one.

4.3 Phenology-Temporal Features Alignment

The details on this module are shown in Fig.4. The temporal encoder outputs K cls tokens, each representing the global temporal features of different tree species. The cosine similarity between each cls token and all other ordinary tokens is calculated, and the results are processed through a softmax activation layer to obtain the Temporal Embedding TE. These embeddings explicitly represents the relationship between the global and local temporal features of different tree species. For the same tree species, the larger the TE value at timestep T, the more significant the local temporal features at that time are for the task of extracting that tree species. The phenological features of tree species refer to the growth and development stages exhibited by trees in different seasons throughout the year. These features typically include

	CF	MP	T	EU	EGU	SBS	MW	SB	FT	ERMF	RHS	All
accarucy	0.7912	0.8386	0.9160	0.9527	0.9383	0.8894	0.8938	0.7755	0.7533	0.8674	0.9377	0.8685/0.9038
iou	0.6869	0.7546	0.8331	0.8378	0.8823	0.7346	0.7079	0.6421	0.6571	0.8272	0.9202	0.7713/0.8245

Table 1. Classification accuracy and IOU of different tree species and all species. The first row of the table takes the first letter of tree species' name. For the accuracy of all species, the left value represents macro accuracy, the right value represents micro accuracy. We apply the same rule to the IOU of all species.

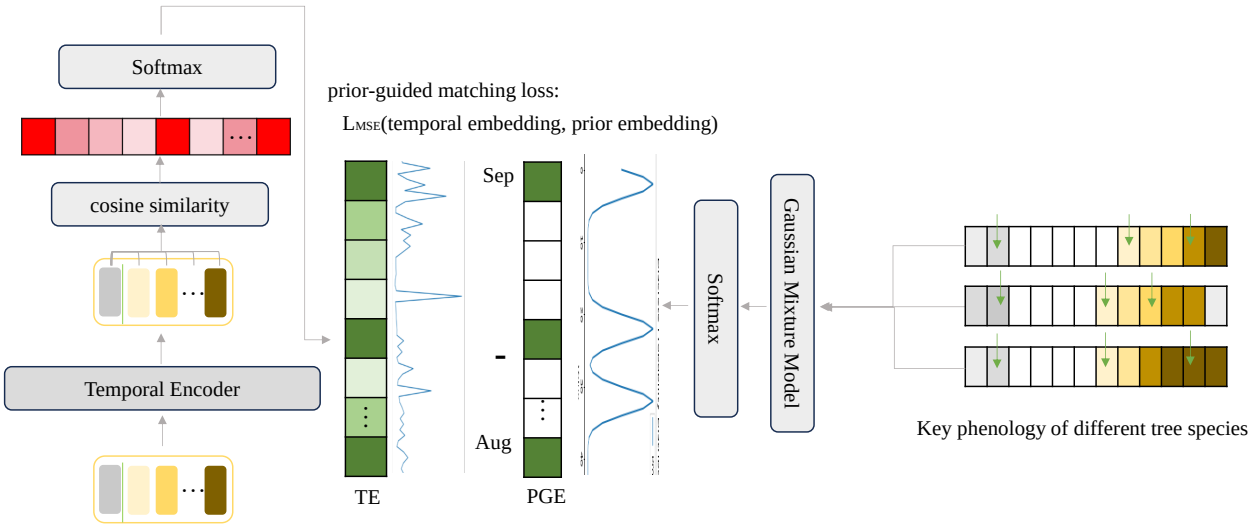

Figure 4. The details on phenology-temporal features alignment module and phenological prior guided loss function. Through T2T and P2T transfer module, we get temporal embedding TE and prior guided embedding PGE respectively.

processes such as budding, flowering, fruiting, and leaf fall. As shown in Fig.4, these features can be mathematically represented as a one-dimensional vector, with the vector values corresponding to specific time points, measured by day of year (DOY). Based on empirical knowledge, the start of the growing season, the start of the maturation season, and the end of the growing season are the time points with the most significant spectral changes for different tree species. Therefore, these three timesteps are selected to construct a one-dimensional vector which represents the parameter μ. This vector is used as a parameter input for a Gaussian Mixture Model (GMM), with the weights w and variance σ set to $1/3$ and 3, respectively. The model ultimately outputs a one-dimensional feature p , as shown in equation(3).

$$p(x) = \sum_{i=1}^{3} w_i \cdot \mathcal{N}\left(x; \mu_i, \sigma^2\right) \tag{3}$$

After normalization, this feature becomes the prior-guided embedding $P\,GE$, which encodes the phenological features specific to tree species observed by ground observation networks. For tree species with missing phenological information, a uniform distribution model is used to construct the $P\,GE$. Now, we have converted both the temporal features and pheno-logical features into one-dimensional features, $T\,E$ and $P\,GE$.

4.4 Integrating Phenological Priors

The model's loss function consists of two components: L_{prior} and L_{seg}, as shown in the equation(4):

$$Loss = L_{seg} + \alpha * L_{prior} \tag{4}$$

where α is a hyperparameter. This loss function uses ground phenological information and tree species label information as supervisory signals to guide the learning of the temporal encoder and spatial encoder.

5. Experiments

5.1 Performance of the model

We evaluate our model on GXData dataset and present the classification accuracies for different tree species (Table.1). We also calculate micro accuracy, micro Intersection over Union(IoU), macro accuracy and macro IoU for all tree species. From the table we observe that the accuracies for all tree species are higher than 0.75, and the IoU values are higher than 0.64, indicating that our model performs well across all species. The confusion matrix visualized in Fig.6 further demonstrates that each tree species is accurately classified.

Soft Broadleaved Species (SBS) and Fruit Trees (FT) exhibit the lowest classification accuracies, which may be influenced by the limited number of labels for these classes. In contrast, Eucalyptus Urophylla (EU) and Eucalyptus Grandis x Urophylla (EGU) achieve the highest accuracies while Rocky Hill Shrub (RHS) and Eucalyptus Grandis x Urophylla (EGU) attain the top two IoU scores. This may be attributed to these

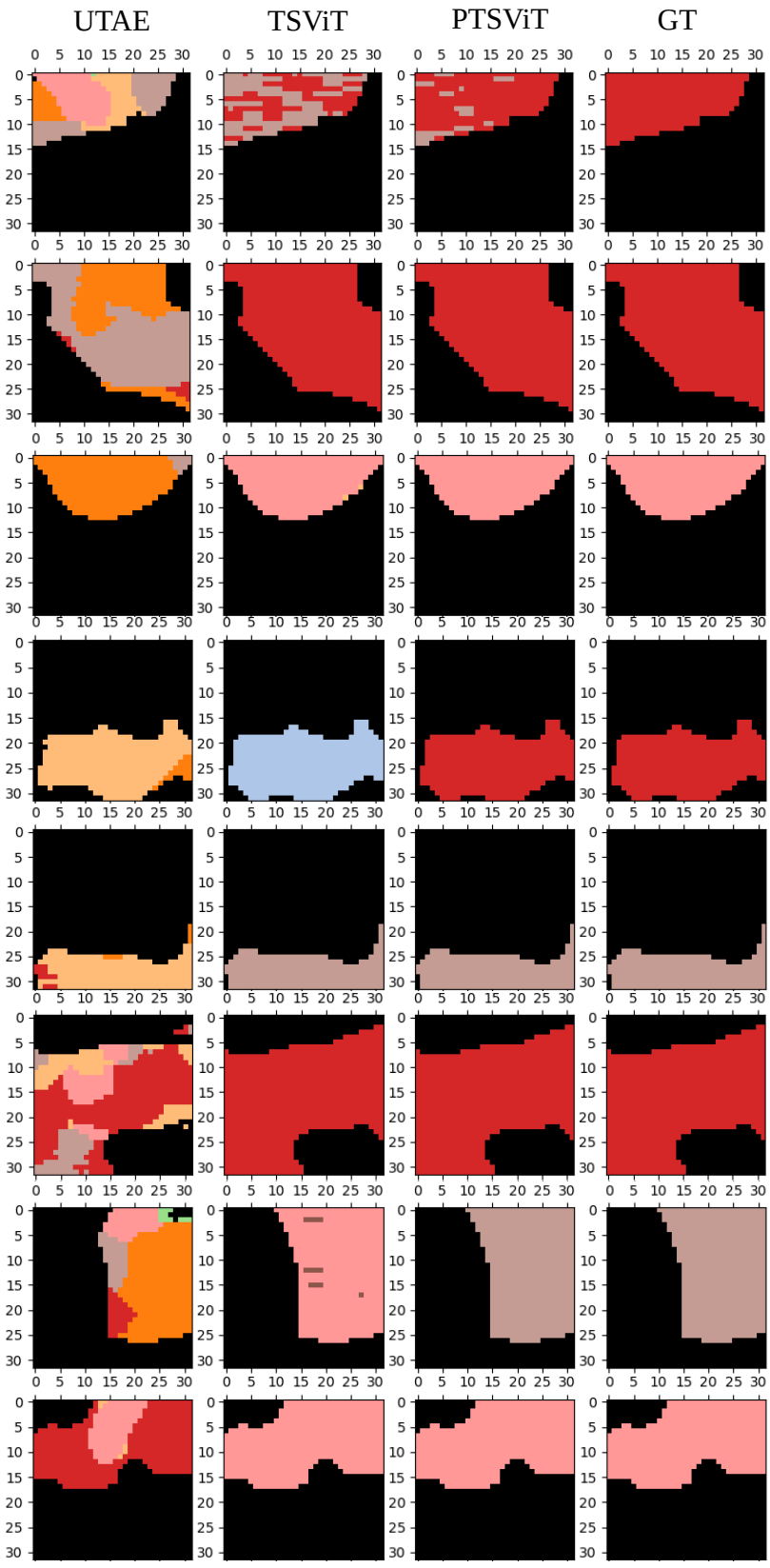

Figure 5. Qualitative examples for GXData.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-559-2025

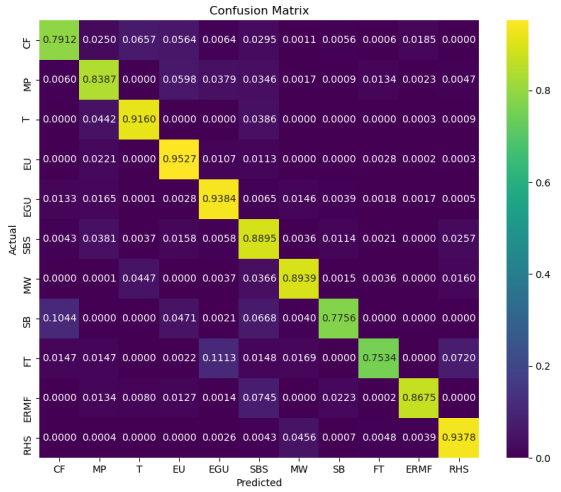

Figure 6. Confusion matrix of PTSViT's result.

species having the most labels; however, since EU and EGU belong to the same genus, the IoU between these two species is lower than that of RHS.

5.2 Ablation study

We conducted ablation study on the designed module using GXData. Table.2 shows that all evaluation metrics improve significantly with the inclusion of this module. Notably, macro IoU improves the most, reaching 77.13% (an increase of +27.53%). Even micro accuracy, which shows the least improvement, increases from 79.10% to 90.38% (+14.26%).

	micro accuracy	micro iou	macro accuracy	macro iou
w/o phenology prior	0.7910	0.6543	0.7406	0.6048
w/ phenology prior	**0.9038**	**0.8245**	**0.8685**	**0.7713**

Table 2. Ablation study for phenology prior guided loss.

The above experiment demonstrates the effectiveness of our model, which integrates phenological priors with deep spatio-temporal features, significantly enhancing the performance of tree species classification. Moreover, the $P2T$ and $T2T$ transfer module successfully transform features from different modalities into the same representation space.

5.3 Compare with other methods

In Table.3 and Fig.5, we present the performance of our final PTSViT model compared to other deep learning-based method on GXData. For all tree species, PTSViT achieves the highest scores across all metrics, improving micro accuracy from 78.72% (achieved by UTAE) to 90.38% (PTSViT). TSViT and UTAE exhibit similar performance, with TSViT performing slightly better.

Fig.5 presents qualitative results of different methods on GXData. UTAE often splits a single parcel into several parts, misclassifying some of them, and occasionally misclassifies entire parcels. In contrast, TSViT, TSViT rarely fragments parcels as UTAE does but sometimes misclassifies whole parcels.

	UTAE	TSViT	PTSViT
micro accuracy	0.7872	0.7910	**0.9038**
micro iou	0.6491	0.6543	**0.8245**
macro accuracy	0.7214	0.7406	**0.8685**
macro iou	0.5437	0.6048	**0.7713**

Table 3. Comparison of different methods for tree species classification. We record four metrics, our method all gets the best result.

Overall, PTSViT demonstrates superior accuracy and robustness compared to both UTAE and TSViT.

6. Conclusion

In this paper, we have proposed PTSViT, a tree species classification model that integrates ground-based phenological information with deep spatio-temporal fusion features. The model comprises a temporal encoder, a spatial encoder-decoder, a phenology-temporal feature alignment module, and a loss function that incorporates phenological priors. We have also created a tree species classification dataset named GXData and demonstrated the effectiveness of incorporating prior knowledge through experiments on this dataset. In future work, we will focus on multimodal representation learning to integrate additional useful information for fine-grained tree species classification.

Acknowledgements

This study was jointly supported by National Key Research and Development Program of China(No.2022YFB3904100, No.2023YFF0725200), the National Natural Science Foundation of China (No.42101446), China Postdoctoral Science Foundation (No.2022T150488).

References

Balestra, M., Chiappini, S., Malinverni, E. S., Galli, A., Marcheggiani, E., 2021. A machine learning approach for mapping forest categories: An application of google earth engine for the case study of monte sant'angelo, central italy. *Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part VII 21*, Springer, 155–168.

Dosovitskiy, A., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Du, Z., Deng, M., Zhang, F., Zhang, M., Chen, J., Guo, J., Li, Q., Qin, L., 2019. Phenological Observations and Analysis of the Cunninghamia lanceolata Seed Orchard in Northern Guangxi(in Chinese). *Modern Agricultural Technology*, 7.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote sensing of environment*, 186, 64–87.

Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1), 185–201.

Garnot, V. S. F., Landrieu, L., 2020. Lightweight temporal self-attention for classifying satellite images time series. *Advanced Analytics and Learning on Temporal Data: 5th ECML PKDD*

Workshop, AALTD 2020, Ghent, Belgium, September 18, 2020, Revised Selected Papers 6, Springer, 171–181.

Garnot, V. S. F., Landrieu, L., 2021. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4872–4881.

Grabska, E., Frantz, D., Ostapowicz, K., 2020. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment*, 251, 112103.

Hemmerling, J., Pflugmacher, D., Hostert, P., 2021. Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sensing of Environment*, 267, 112743.

Hermosilla, T., Bastyr, A., Coops, N. C., White, J. C., Wulder, M. A., 2022. Mapping the presence and distribution of tree species in Canada’s forested ecosystems. *Remote Sensing of Environment*, 282, 113276.

Houghton, R., 2005. Aboveground forest biomass and the global carbon balance. *Global change biology*, 11(6), 945–958.

Kollert, A., Bremer, M., Löw, M., Rutzinger, M., 2021. Exploring the potential of land surface phenology and seasonal cloud free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region. *International Journal of Applied Earth Observation and Geoinformation*, 94, 102208.

Kanan, A.H., Pirotti, F., Masiero, M., Rahman, M.M., 2023. Mapping inundation from sea level rise and its interaction with land cover in the Sundarbans mangrove forest. *Climatic Change* 176, 104. https://doi.org/10.1007/s10584-023-03574-5

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5), 823–870.

Martinez, J. A. C., La Rosa, L. E. C., Feitosa, R. Q., Sanches, I. D., Happ, P. N., 2021. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171, 188–201.

Qin, G., 1997. *Cultivation and Utilization of Masson Pine(in Chinese)*. Jindun Press.

Rußwurm, M., Korner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7(4), 129.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28.

Tarasiou, M., Chavez, E., Zafeiriou, S., 2023. Vits for sits: Vision transformers for satellite image time series. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10418–10428.

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., Roy, D. P., 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, 185, 271–283.