Monitoring Spartina Alterniflora Using Self-Supervised Learning

Othmane Echchabi, Advait Bhaskar Pandit, Bayan Hameed Alharbi, Meixiang Du
Project Manager: Keqi He Project Lead: Wenhong Li, Ding Ma

Duke + Programs

Introduction

Spartina alterniflora

a native species vital to the health of Atlantic coast wetlands in the United States, is facing significant threats from climate change, which jeopardize its crucial role in carbon sequestration, water quality maintenance, and habitat provision.

Concurrently, *Spartina alterniflora* has become an invasive species along the coastal regions of China since its introduction in 1979 for shoreline stabilization. This has led to substantial ecological consequences, disrupting native ecosystems and biodiversity.

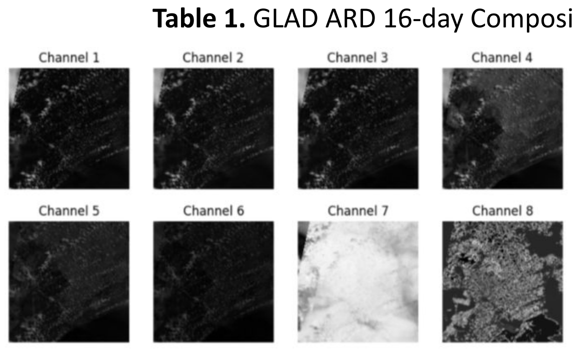
Our study employs satellite imagery and self-supervised learning models to map and monitor the distribution of *Spartina alterniflora* with minimal in-situ data. This approach addresses the challenges of sparse labeled data, offering a cost-effective and scalable solution for ecological monitoring. By understanding the spatiotemporal dynamics of this species, we aim to inform conservation strategies and enhance the resilience of coastal wetlands in the face of climate change and biological.

Data

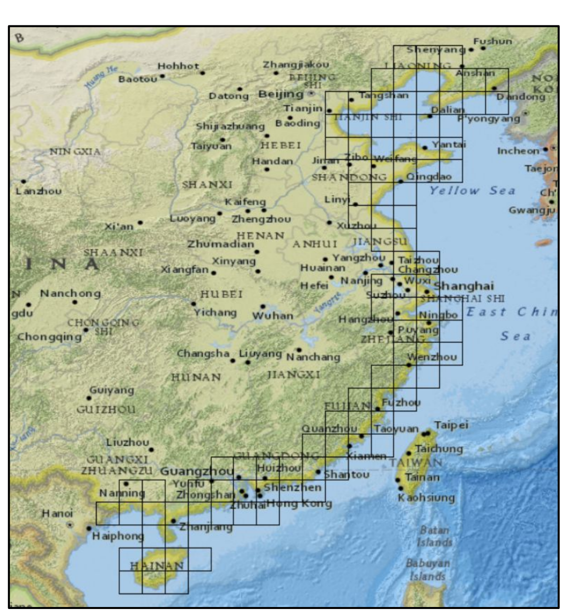
GLAD ARD Dataset: We used a 16-day time series of globally consistent, tiled Landsat normalized surface reflectance satellite images from 1997 to the present. Each 16-day interval for a tile is stored as 8-band, 16-bit unsigned, LZW-compressed GeoTIFF file. Each of the 8 bands contains information from varying wavelengths of light as shown by the chart below.

Band	Image data	Units, data format
1	Blue band (0.48 µm)	Normalized surface reflectance scaled to the range from 1 to 40,000, UInt16
2	Green band (0.56 µm)	
3	Red band (0.66 µm)	
4	NIR band (0.86 µm)	
5	SWIR1 band (1.61 µm)	
6	SWIR2 band (2.20 µm)	
7	Normalized brightness temperature	K × 100, UInt16
8	Observation quality flag (QF)*	QF code, UInt16

Table 1. GLAD ARD 16-day Composite Data Format



Tile Selection: We selected tiles from US and China with high density of Spartina Alterniflora



Data Pre-Processing: To prepare the data for the learning process, we employed a semi-random series of transformations: scaling, cropping, skewing, and hue changes.

We also used the Google Earth Explorer API to download Sentinel-2 and NAIP images for they respective high resolutions of 10 meters per pixel and 0.6 meters per pixel.

GEE Pipeline github repository:
https://github.com/othmaneechc/gee-exporter
Website: https://www.savecoordinates.com/

Methods

First Method: SimCLR (CNN–Object-based):

Pretext task: Define pretext task through stochastic data augmentation (cropping, flipping, colorization, Gaussian blur, affine transformation). Employ ResNet-50 as the base encoder to extract representation vectors from augmented data examples. Build a training pipeline by optimizing the model's performance by minimizing the contrastive loss of contrastive prediction tasks.

Downstream task: Conduct knowledge distillation by freezing the representations learned from the pretext task. Incorporate the representations into the U-Net model structure to enable segmentation in the downstream task.

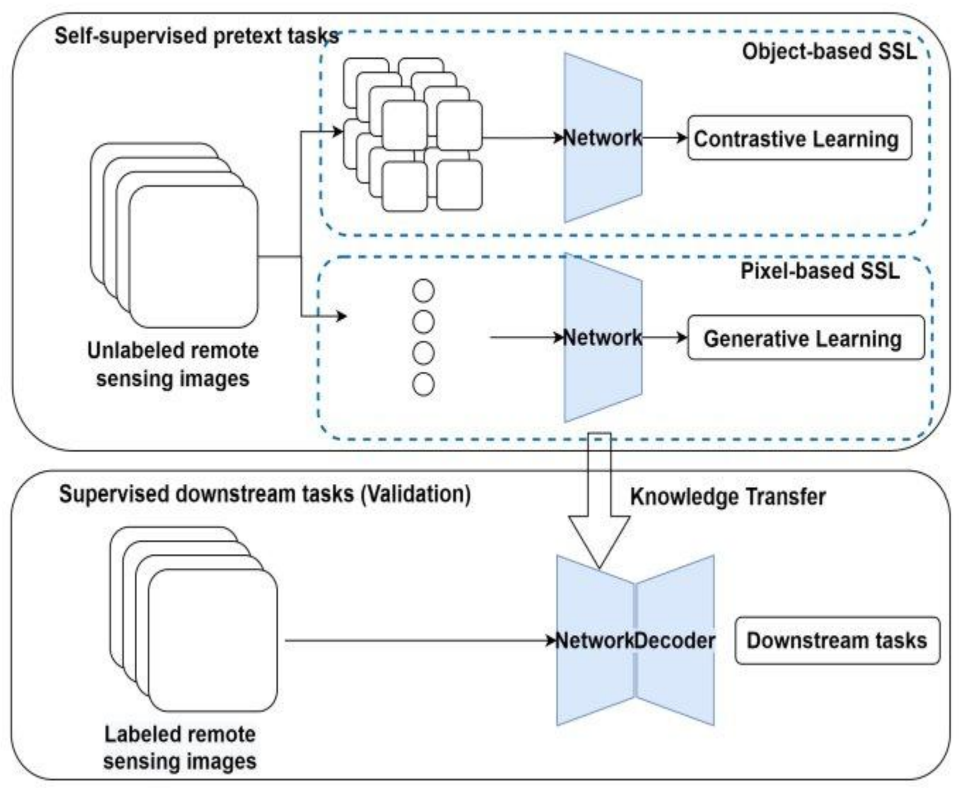
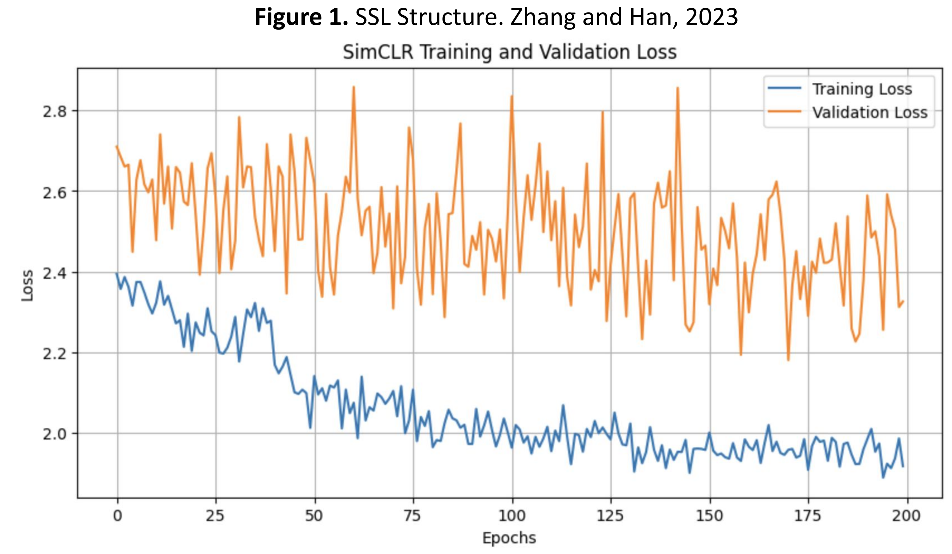
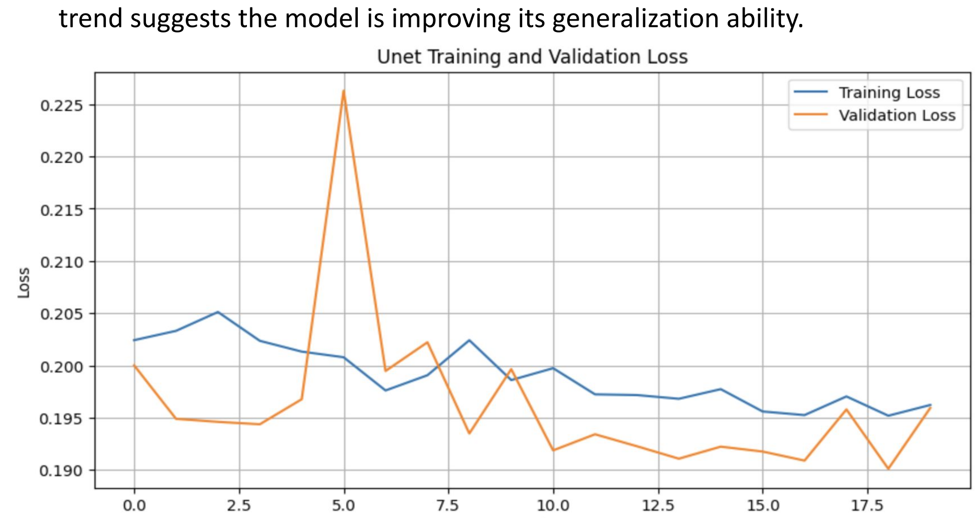


Figure 1. SSL Structure. Zhang and Han, 2023



Graph 1. SimCLR Training. Due to absence of annotated Spartina data, We used VOC Segmentation, a subset of the PASCAL Visual Object Classes (VOC) dataset to test downstream pipeline.

Observation:
The training loss decreases as the number of epochs increases, indicating that the model is learning, yet the the pace is slow due to the potential discrepancy between Spartina data and the VOC data.
The validation losses fluctuate due to data sparsity, but an overall downward trend suggests the model is improving its generalization ability.



Graph 2. Unet Training on 300 real Landsat-9 satellite images for training and 23 images for validation.

Observation:
The training loss steadily decreases as the number of epochs increases, indicating that the model is learning effectively.
The validation losses fluctuate due to data sparsity, but an overall downward trend suggests the model is improving its generalization ability.

Second Method: DINO (Self-Supervised Vision Transformer)

Pretext Task: Define the pretext task using self-supervised learning with Vision Transformers (ViT). Apply stochastic data augmentation techniques (cropping, flipping, color jittering, and random erasing). Use a dual-branch architecture, one branch processes the original images, the other processes augmented views. Employ a teacher-student setup: the teacher model is an exponential moving average (EMA) of the student model. Optimize the model by minimizing the discrepancy between the teacher and student outputs using a mean squared error (MSE) loss.

Attention Visualization: Utilize the attention heads from the trained DINO model to visualize the learned features. Investigate if the attention heads can effectively detect and highlight Spartina alterniflora in the input images.

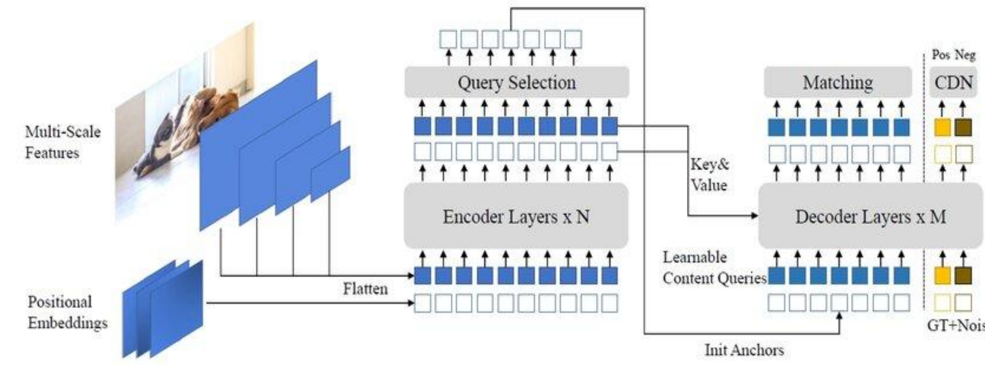
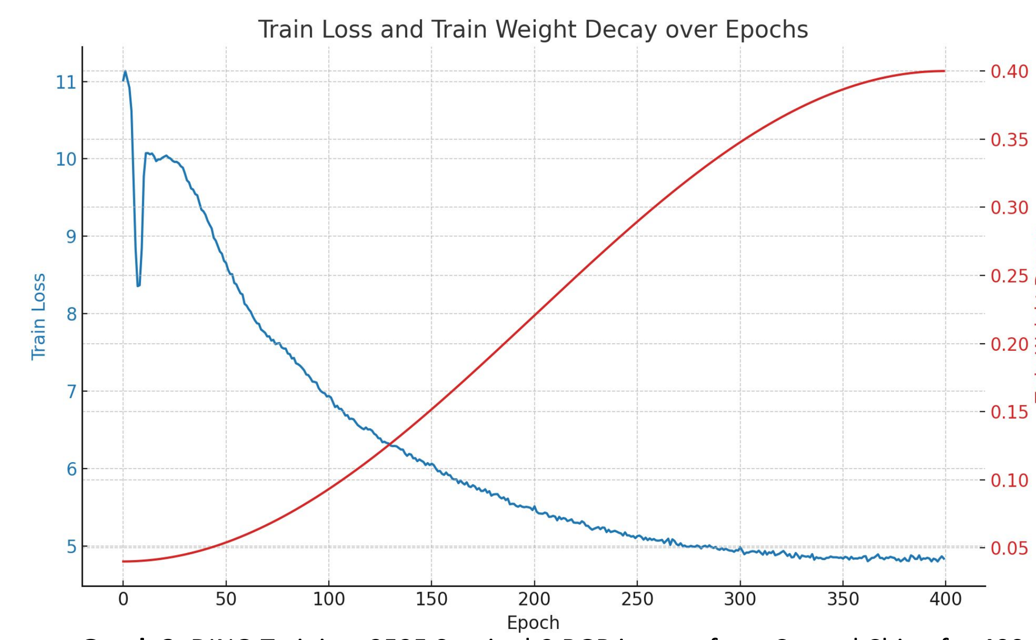
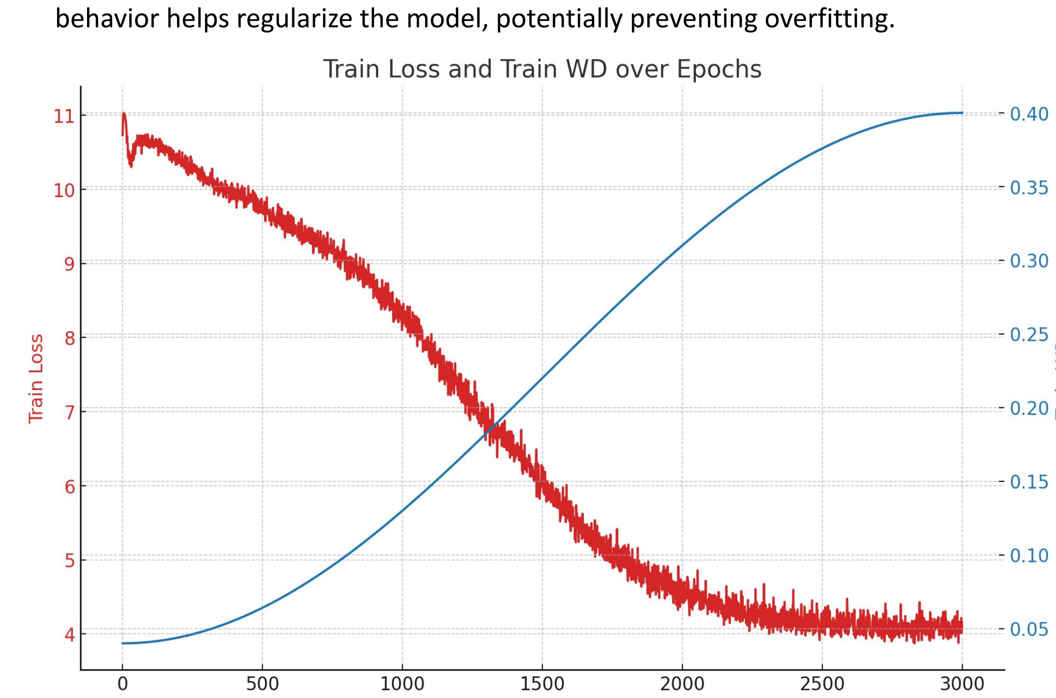


Figure 2. Distillation with No labels **(**DINO**)** Model Structure.



Graph 3. DINO Training. 2525 Sentinel-2 RGB images from Coastal China for 400 epochs.

Observations:
Training Loss: The training loss consistently decreases, effectively learning.
Weight Decay: The train weight decay shows a steady increase over epochs. This behavior helps regularize the model, potentially preventing overfitting.



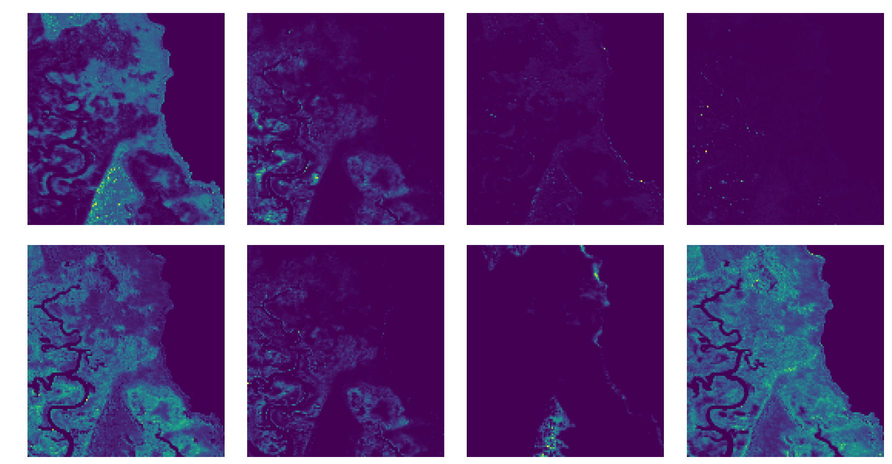
Graph 4. DINO Training. 286 NAIP IR images from Coastal Virginia for 3000 epochs.

Observations:
Training Loss: The training loss exhibits a decreasing trend as the number of epochs increases.
Weight Decay: The training weight decay shows a steady increase over the epochs.
Overall Trend: The combined plot illustrates a significant relationship between the model's learning process and its regularization. As the training loss decreases, indicating improved learning, the increasing weight decay ensures that the model remains generalized and robust.

Preliminary Results

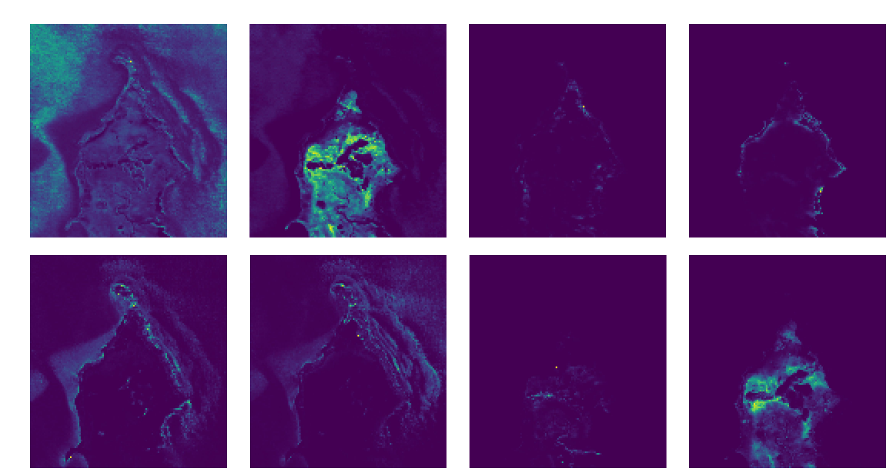
We utilized a base Vision Transformer (ViT) model that outputs 12 attention heads. For our analysis, we retained 8 of these attention heads and discarded those that focused on features in the ocean, as our goal is to identify Spartina alterniflora. The images presented here are from Coastal Virginia and include Near Infrared (NIR), Red, and Green (IR) bands.

NAIP Image: The image displays a coastal region of Virginia with prominent vegetative features highlighted in red due to the NIR band. The winding waterways and coastal features are clearly visible.
We ended up using NAIP Satelitte images because they have a high resolution of 0.6 meters per pixel



Attention Heads: The subsequent images represent the attention maps from the ViT model. Each attention head focuses on different aspects of the image. Some heads highlight the vegetative areas strongly, others capture the water bodies and their boundaries few heads seem to identify the transitions between land and water effectively.

The combination of NIR, Red, and Green bands in the NAIP images enhances our ability to detect and monitor vegetation in coastal regions of Virginia. There is significant potential in improving the detection of Spartina alterniflora by:



Future Work

Using More Bands: Incorporating additional spectral bands can provide more detailed information, further improving the accuracy of vegetation detection. The self-supervised model SatMAE takes multiple bands.

Labeled Data: Acquiring labeled data for training can significantly enhance the model's performance by providing ground truth references, allowing the model to learn more effectively and accurately identify Spartina alterniflora.