Ghost Forest Detection with Sub-meter Resolution Satellite Imagery

April 10, 2025

Stakeholder Names and Roles

Stakeholder	Role
Professor Xi Yang	*Sponsor: UVA Environmental Sciences Professor*
Henry Yeung	*Sponsor: UVA Environmental Sciences Graduate Student*
Professor Heman Shakeri	*Mentor*
Mahin Ganesan	*Member*
Brendan Jalali	*Member*
Tom Lever	*Member*
Nicholas Miller	*Member*

Abstract

We are interested in identifying ghost forests across the United States. A ghost forest is a region with more than 10 dead trees per hectare along the coast. The objective of our project is to dive deep into applying deep learning algorithms to image detection in order to be able to map individual dead trees. To accomplish this goal, we have developed, using an existing neural network for urban tree detection (Ventura et al.), a neural network that predicts the centroids of dead trees in test images with an F1 score of 0.77. We hope to improve this neural network and allow it to predict geospatial layers not only of centroids of dead trees but also probabilities that pixels represent dead trees. We will deliver code for the neural network and for generating validation and testing performance metrics, annotated testing images, and geospatial layers based on inference images. We have created training, validation, and testing data sets based on satellite images of vulnerable lowland areas and annotation images. We have also tested other model architectures, different loss functions, and data augmentation in hopes of boosting model performance. Our data include satellite images that depict images a few hectares in area near the East Coast. Overall, the goal of this project is for the members to participate in developing a data science pipeline to identify ghost forests, which will in turn provide us an understanding of where and why ghost forest hotspots may arise. The locations of ghost forests are presently unknown. Knowing where ghost forests are occurring, governments and localities can push for legislation to keep these forests from dying, allowing them to continue protecting the surrounding coastal communities and ecosystems.

Outline of the Project

For the study and preservation of coastal forests, our sponsors, who are researchers in the Department of Environmental Science at the University of Virginia, have taken an interest in developing a tool that will be able to identify the extent of ghost forests. Our sponsors are interested in looking at the creation mechanisms of ghost forests, and in being able to identify more locations. We are taking the first step towards coastal forest preservation. In order to satisfy these requests, we have developed a model that predicts the location of dead trees in ghost forests with an F1 score of 0.77. Our model is valuable in predicting the locations of ghost forests across the United States. With this improved model, this effort will lead to a better understand the systems that cause the death of these forests. We are assuming high quality information from satellite imagery in order to build our model. Our scope is to build a neural network that, by the end of this semester, will identify coastal ghost forests in the United States that have potential for saltwater intrusion.

Success Criteria

Our success criteria are defined in the following table.

SC1	*Identify potential ghost forest locations that can be used to expand training and testing*
SC2	*Build data pipeline for initial supervised model*
SC3	*Build and train supervised model that successfully meets expectations*
SC4	*Ensure that model predicts heat map of probabilities that pixels represent dead trees*
SC5	*Improve architecture of model for better performance in predicting dead trees*
SC6	*Passing the capability onto our sponsors, including documentation and Slurm script for Model execution*

Green demarks completion

Data Assumptions and Limitations

The data we are using currently for training our supervised model is data provided by our sponsors that include satellite images from the East Coast of the United States. Along with this data, we received annotated images to accompany these satellite images. We have come across some limitations expanding this data set with satellite images of locations along the Gulf Coast, West Coast, and Great Lakes. We are searching for professionals who can create annotation images for new satellite images and accelerate expanding our data set. Our first task was preparing to expand the data set from already gathered East Coast data points to include the Gulf Coast, West Coast, and Great Lakes Coasts. We prepared by identifying hundreds of locations with dead trees. After preparing to expand, we moved forward with the East Coast data provided by our sponsors. If we expand beyond the East Coast data that we are working with currently, we may run into limitations with data storage, transfer, processing time, and time to run a neural network.

We are relying on annotated images correctly encoding polygons that circumscribe dead trees. We note that training images may consist of multiple 256 x 256 tiles going both left to right and top to bottom. Training and validation images are fully annotated. On the other hand, only central 166 x 166 sections of training images are annotated. Inference images are not annotated.

As of 4/10/2025, we have progressed steadily and succeeded in:

- identifying potential ghost forest locations that can be used to expand training, validation, testing, and inference data sets
- Based on satellite images and annotation that our sponsor provided, creating 256 x 256 image tiles and tables of coordinates of centroids of dead trees that save as data for a neural network
- Using Rivanna GPUs, training, validating, testing, and inferring with a neural network
- Fixing and augmenting neural network
- Experimenting with ResNet
- Preparing to experiment with other loss functions
- Preparing to experiment with image augmentation

Summary of Data Processing and Data Aggregation

Our data processing pipeline begins with the data provided by our sponsors, whose source is the NAIP (National Agriculture Imagery Program). We have pulled that data into Rivanna and have built a system that takes greyscale images representing red, green, blue, and near infrared channels of satellite images, and their corresponding annotation images. We create RGBN images based on the channels, and corresponding data tables of centroids of coordinates of dead trees. We tile images and tables of centroids as required by the neural network. These pairs of tiled images and data tables of marked dead trees then serve as the input for the training, validation, and testing of our model.

We ultimately intend to predict not only geospatial layers with centroids of dead trees but also heat maps of probabilities that pixels in satellite images represent dead trees. For training, validation, and testing, our model ingests 256 x 256 tiles of satellite images and tables of coordinates of centroids of dead trees. For inference, our model ingests satellite images with arbitrary dimensions. For training and validation, our model outputs logs and performance metrics. For testing, our model outputs logs, performance metrics, and annotated satellite images. For inference, our model outputs logs and a JSON file that can be laid over a satellite image in QGIS and used to visualize centroids of dead trees.

Data Visualizations

Figure 1: Tile-able Satellite Image	Figure 2: Qualitative, Testing, False Color Visualization w/ Predicted Annotations

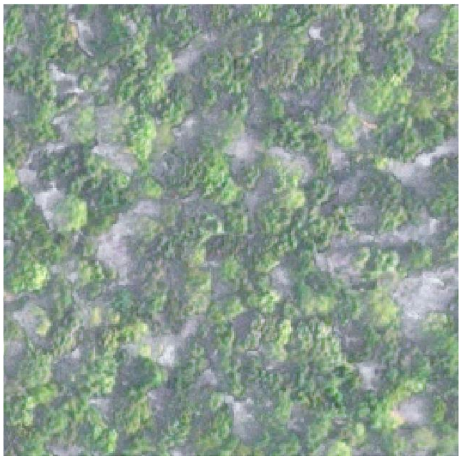
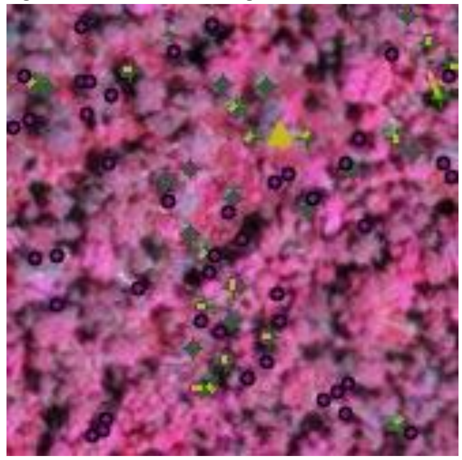
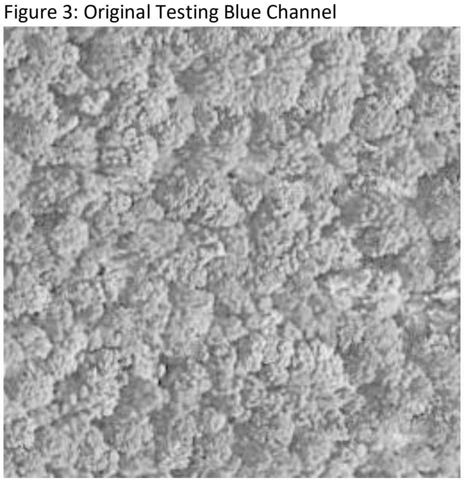
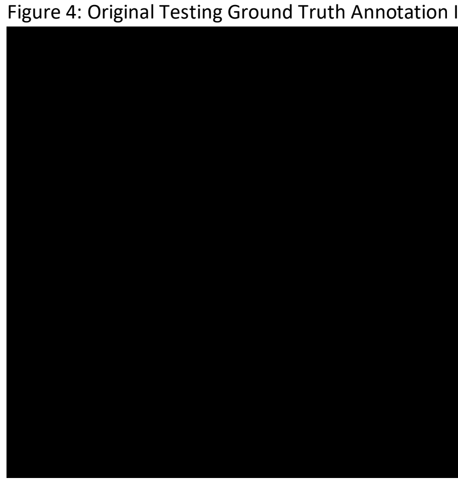

	

Figure 3: Original Testing Blue Channel	Figure 4: Original Testing Ground Truth Annotation Image
	

Summary of Modeling and Analysis

Our sponsors directed us to adapt a neural network that detects trees in urban environments to detect dead trees in satellite images. The article by Ventura et al. published in 2024 called, “Individual Tree Detection in Large-Scale Urban Environment using High-Resolution Multispectral Imagery” describes the architecture of the neural network that we adapted. We have processed data to be compatible with this neural network. We intend to ensure that our model predicts not only geospatial layers of centroids of dead trees but also heat maps of probabilities that pixels represent dead trees and to continue to improve our architecture. The baseline architecture can be described below in Figure 1.

Figure 5: Architecture	Figure 6: Loss Functions

Appendix A. HR-SFANet architecture

Here we provide pseudo-code to exactly describe the architecture of our HR-SFANet network, which is a modification of the original SFANet architecture Zhu et al. (2019). The pseudocode makes use of the following standard neural network operations:

- Conv(n, k, x): 2D convolution on input x with a filter size of $k \times k$ and n output channels.
- ReLU(x) = max($0, x$): Element-wise rectified linear unit.
- Sigmoid(x) = $1/(1 + \exp(-x))$: Element-wise sigmoid function.
- BatchNorm(x): Batch normalization (Ioffe and Szegedy, 2015).
- MaxPool(x): 2x2 max pooling with a stride of 2.
- Concatenate(x, y): Channel-wise concatenation of x and y.
- Upsample(x): $2\times$ upsampling of input x using bilinear interpolation.

$$L_{\mathrm{MSE}} = \frac{1}{N}\sum_{x,y}[C'(x,y) - C(x,y)]^2$$

$$L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{x,y}[A(x,y)\log(A'(x,y)) + (1 - A(x,y))\log(1 - A'(x,y))]$$

$$L = L_{\mathrm{MSE}} + \alpha L_{\mathrm{BCE}}$$

Let's consider the neural network architecture in depth. A batch of data for our model consists of a 256 x 256 satellite image with red, blue, green, and alpha channels, and an annotation image with non-black pixels representing vertices of polygons circumscribing dead trees. Our model uses 2D convolutional layers to detect features of dead trees. The model also uses pooling layers to look at specific areas of an image at a time and combine portions of the image to reduce the dimensions, ensuring focus on the most important features. Finally, the model adds an attention map to focus on regions where dead trees are likely to be located and a confidence map to show the likelihood of dead trees at any given pixel.

In this architecture, the loss function is broken down into two parts, MSE loss and binary cross entropy loss. MSE is the main loss function and compares predicted and actual ground truth confidence maps. The ground truth map includes indicators that pixels represent centroids of dead trees. The ground truth map is compared to the model's predicted map of probabilities whose pixels represent centroids of dead trees. MSE loss is used to see how much the model differs from the ground truth when searching for dead trees in training and validation images. The binary cross entropy loss focuses on the attention map, seeing how well the predicted regions of dead trees match the actual locations of dead trees. Our final loss calculation adds the MSE loss to a weighted BCE loss (depending on how important we find each loss to be) to get our total loss function as seen in Figure 2 above.

We use the architecture of Ventura et al., and we have applied our knowledge of Deep Learning to improve this architecture and our neural network's prediction of centroids of dead trees. We intend to predict heat maps of probabilities of pixels representing dead trees. We have evaluated the performance of our model and have made improvements to our model and its predictions, by examining loss curves derived during training and validation, and precision, recall, F1 score, and Root Mean Squared Error calculated during validation and testing. Below are some of our intermediary performance metrics:

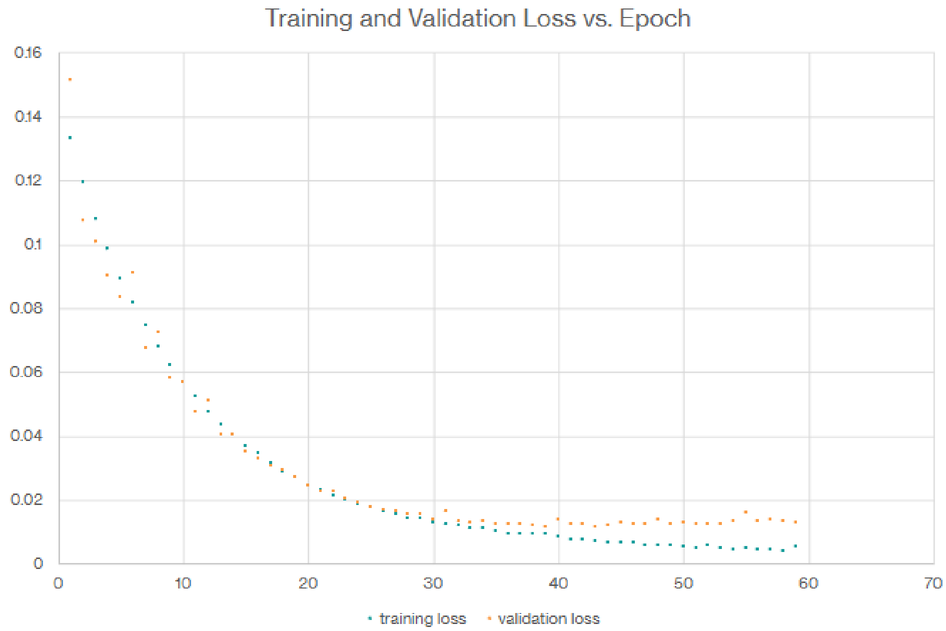


Figure 7

Metric	Validation	Testing
Precision	0.697	0.784
Recall	0.676	0.747
F-Score	0.686	0.765
RMSE [px]	2.714	2.693

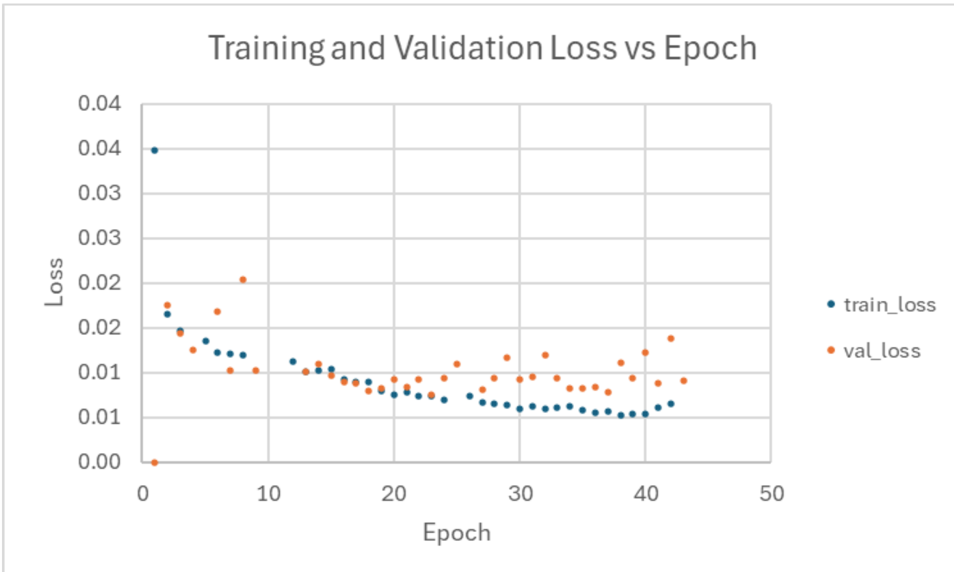
Our primary performance metrics of the improved model are precision, recall, F1 score, and RMSE. Precision measures the number of true positives divided by the number of total positives in the model, both true and false. In the scope of our project, precision represents how many of the predicted dead trees are actually dead trees. Our intermediary precision of about 78% indicates that about 78% of the model's predicted dead trees were indeed dead according to ground truth. Recall indicates our model's ability to capture as many dead trees as possible, for which we were able to get 75% using our intermediary model. Our F1 score of 77% represents the performance of the model when combining both precision and recall. We have goals of 80% precision, F1 score, and recall. Finally, we are using RMSE to compare the distances between true positive predictions of dead trees and the ground truth map counterparts. We currently have an RMSE of 2.693. A lower RMSE in future experiments would indicate that the model's predicted dead tree centroids are closer to the dead tree ground truth locations.

In the future, we plan to improve the results of this model by changing this model, adjusting hyperparameters, trying multiple loss functions, and trying data augmentation.

ResNet Architecture Testing

To try improving our results, we swapped out Ventura's VGG backbone with ResNet50, a deeper neural network that theoretically can better map complex relationships. The model also alleviates the vanishing gradients problem with its novel skip connection idea, making it one of the premier deep CNNs.

Below is a loss curve from training the model on our data set, which converges faster than the VGG model:



While ResNet50 converges faster than VGG, its test performance results are worse:

Metric	Testing
Precision	0.67
Recall	0.63
F-Score	0.65
RMSE [px]	2.61

In the future, we plan to improve the results of this model by adjusting hyperparameters, trying multiple loss functions, and using data augmentation. In addition, we plan to try EfficientNet, another model approach, and get a better understanding of why VGG performed better than ResNet to see how we can improve ResNet performance.

Data Augmentation

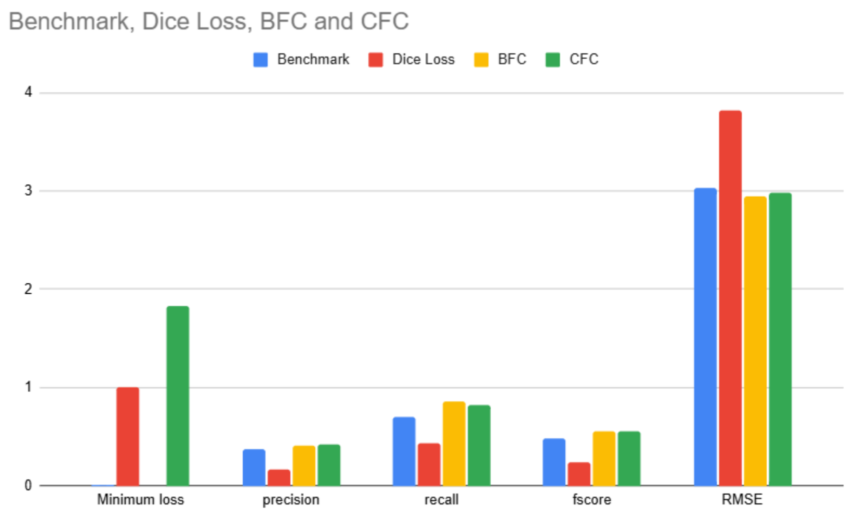
In the implementation of Data augmentation, we were able to expand the scope of the data set. We performed several transformations to the images, specifically 90 degree rotations, and vertical flips of those rotations. This allowed us to expand the data set by a factor of 8, which was great for the scale of our training and testing sets. This expanded our preparing, training, tuning, and testing time, but allowed for the improvement of the model performance. We will continue to test and scale this approach. Additionally, in order for this project to be handed off successfully to our sponsor, we will implement a SLURM script that will allow for easy use of Rivanna's resources. This will also include work with improving the variety of data augmentation techniques such as random cropping and brightness and contrast changes.

Loss Functions

As previously stated, our current loss function is composed of two terms: Mean Squared Error (MSE) for the confidence map and Binary Cross Entropy (BCE) for the attention map. We also explored alternative loss formulations, including the Dice loss, which is commonly employed in segmentation tasks due to its sensitivity to spatial overlap between predicted and ground truth masks. Equations used for the loss functions can be seen below:

Loss Function	
Binary Cross Entropy	$\mathrm{BCE}(y, \hat{y}) = -\left[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})\right]$
Dice Loss	$\text{Dice Loss}(y, \hat{y}) = 1 - \frac{2 \sum y\hat{y} + \epsilon}{\sum y + \sum \hat{y} + \epsilon}$
Binary Focal Cross Entropy	$\mathrm{BFCE}(y, \hat{y}) = -\alpha \cdot (1 - \hat{y})^{\gamma} \cdot y \cdot \log(\hat{y}) - (1 - \alpha) \cdot \hat{y}^{\gamma} \cdot (1 - y) \cdot \log(1 - \hat{y})$
Categorical Focal Cross Entropy	$\mathrm{CFCE}(y, \hat{y}) = -\sum_{c=1}^{C} y_c \cdot (1 - \hat{y}_c)^{\gamma} \cdot \log(\hat{y}_c)$

To rigorously assess the effectiveness of these loss functions, we compared performance across several variants. The graph below summarizes the minimum loss achieved during training, as well as key evaluation metrics — precision, recall, F1 score, and Root Mean Squared Error (RMSE) on the test set.



The Benchmark configuration, a combination of MSE and BCE losses, strikes a moderate balance between recall and precision, yielding a respectable F1 score of 0.48. However, it suffers from relatively high RMSE, suggesting less spatial accuracy in centroid prediction. The Dice Loss variant underperforms in all metrics, indicating that while Dice loss is advantageous in segmentation settings with large contiguous masks, it may not effectively optimize centroid-level prediction tasks involving sparse annotations.

Interestingly, the BFC (Binary Focal Confidence) and CFC (Combined Focal Confidence) variants demonstrate improved performance across most dimensions. These losses emphasize the difficulty of classifying pixels, a characteristic particularly useful in our imbalanced context where true positive dead tree centroids are sparse relative to the background. Both BFC and CFC achieve higher precision and recall, with an F1 score of 0.56, indicating more effective trade-offs between false positives and false negatives. Furthermore, both variants reduce RMSE relative to the baseline, signifying better spatial fidelity in centroid localization.

Future Work Plan

Task Description	Date
Along with geospatial layers of centroids of dead, predict heat maps of probabilities of pixels representing dead trees.	End of semester
Achieve better performance metrics by changing model, introducing image augmentation, trying other loss functions, and finding better hyperparameters.	Throughout semester, completed by end of Semester
Create an Architecture Diagram.	End of Semester

Potential Concerns [C] and Blockers [B]

Identifier	Description
C	Data Breadth: Do we have enough data?
B	Will we receive annotation images for satellite images of areas near the Gulf Coast, West Coast, and Great Lakes?

Questions for Reflection

Think back on your capstone experience and discuss these questions as a team. You may include perspectives from the different team members where appropriate.

Reflection Question 1: What was the biggest challenge that you faced with this project?

This project has laid out many challenges but a key challenge that we have been working on and continue to work on is the adaptation of the repository that we had originally formed our model on and understanding both the underlying data and the approach behind our pipeline. We also ran into difficulties with implementation of improvements to the model but specifically it was difficult to apply the theory we have learned throughout the program to real research.

Reflection Question 2: Did this project stretch you to grow? If so, how?

This project has helped each of us stretch and grow by allowing us to work through a model implementation that was extremely technical project, while also handling interpersonal and contribution expectations.

Reflection Question 3: Do you believe the capstone experience will be helpful for your career? If so, how?

We have gained significant experience with a full, professional data science pipeline and collaborating as a team, something that is commonplace in a corporate environment. This project helped us understand the scope of working with real data as well as the hurdles and roadblocks that come with data wrangling and model building.

Reflection Question 4: Anything else that you would like to share?

It is really encouraging to see the progress that we have made, and we are excited to see our final model performance. It will also be exciting to be a part of a larger goal of improving data science and to learn of the impact that this will have on the research that our sponsors are doing.