Identifying Dead Trees in High Resolution Satellite Imagery with Various Foundation Models

Sarah King ([asking@ucdavis.edu](mailto:asking@ucdavis.edu))

Contents

[Motivation: 2](#_Toc174546938)

[Sample Area: 2](#_Toc174546939)

[Methodology: 2](#_Toc174546940)

[Creating labeled masks: 2](#_Toc174546941)

[Creating dataset: 2](#_Toc174546942)

[Training models: 2](#_Toc174546943)

[Results Part 1: Model Comparison 3](#_Toc174546944)

[Broad Conclusions: 7](#_Toc174546945)

# Motivation:

Automate location of Alaskan trees killed by spruce bark beetle damage.

# Sample Area:

Kenai peninsula and area near Nancy Lake State Recreation Area in Alaska



# Methodology:

## Creating labeled masks:

Images were processed using the Gray Level Co-occurrence Matrix (GLCM) augmented clustering algorithm from the senegal-lcluc-tensorflow project. GLCM runs on the red band.

## Creating dataset:

A set of 7 images and labeled training masks were used to create a training dataset (containing 5 images), a validation dataset (containing 1 image) and a test dataset (containing 1 image). Each image was tiled into 256x256 pixel patches with the patchify package, then empty tiles (tiles only containing 0 or 1 values) were removed.

## Training models:

SAM and DINOv2 models were trained with transfer learning on a training dataset size of 10, 100, 500, 1000, and 5000 images. Results are listed in the results section.

As a benchmark, 3 older CNN-based models were likewise trained with varying dataset sizes for comparison. These models include:

ResNet152 + ImageNet Backbone. A 152-layered model with initial model weights from ImageNet pre-training. Final layers were modified for binary prediction task.

ResNet152 (no pretraining). Same model as prior, but without initialized weights from ImageNet pre-training.

A 53 layered CNN structured as follows:

Initial preprocessing block (Initial convolution + Batch normalization + ReLU + pooling layer)

16 blocks (each containing 3 convolutional layers) > Final convolution + interpolation layer for binary prediction.

# Results Part 1: Model Comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Comparing Total Model Accuracy** | | | | | |
| **Training set size** | **SAM** | **DINOv2** | **ResNet152 (ImageNet)** | **ResNet152 (None)** | **CNN** |
| 10 | 0.691 | 0.965 | 0.932 | 0.704 | 0.841 |
| 100 | 0.901 | 0.974 | 0.895 | 0.959 | 0.945 |
| 500 | 0.858 | 0.974 | 0.956 | 0.967 | 0.962 |
| 1000 | 0.856 | 0.974 | 0.960 | 0.964 | 0.970 |
| 5000 | 0.849 | 0.975 | 0.900 | 0.971 | 0.970 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Comparing Model IoU** | | | | | |
| **Training set size** | **SAM** | **DINOv2** | **ResNet152 (ImageNet)** | **ResNet152 (None)** | **CNN** |
| 10 | 0.0294 | 0.4717 | 0.0000 | 0.1406 | 0.0274 |
| 100 | 0.0437 | 0.5885 | 0.2023 | 0.3476 | 0.3776 |
| 500 | 0.0856 | 0.5471 | 0.3661 | 0.4618 | 0.3864 |
| 1000 | 0.1434 | 0.5566 | 0.4217 | 0.4362 | 0.4482 |
| 5000 | 0.1549 | 0.5970 | 0.3448 | 0.5253 | 0.5319 |

A picture containing text, stationary, pencil

Description automatically generated



A picture containing graphical user interface

Description automatically generated

# Broad Conclusions:

Plain language:

* SAM has the worst performance across most metrics.
* Traditional CNN models (eg. ResNet or our custom CNN) do better with more data and outperform SAM.
* DINOv2 has the best performance by most metrics, and notably is able to reach and maintain a high level of accuracy even with smaller amounts of data (in this specific use case, with datasets = 100 images).

DINOv2 is overall the better model based on the metrics of variation in accuracy, total accuracy, IoU, and amount of true positives and negatives found relative to false positives and negatives.

Notably, when looking at metrics of total accuracy and mean IoU, DINOv2 converges near peak performance fairly early, at a training set size of 100 images, where total accuracy hovers around 0.974, mean IoU ranges between 0.5471 and 0.5970, and variation in accuracy remains consistent as well.

On the other hand, SAM seems to perform increasingly better with larger datasets by metrics of IoU and the percentage of true positives found.

SAM’s performance on a dataset size of 5000 is comparable to ResNet-152’s performance on a dataset of the same size. However, ResNet-152 outperforms by metric of mean IoU and percentage of true positives and negatives found.