

Recommendations for The Whitney Lab

Fossil Computer Vision Modeling

Adrian Reuschel, Leah Boger, Kevin Rabinak

2025-12-05

This document was prepared for Dr. Megan Whitney by graduate students in STAT495: Data Science Consulting. Students worked with lab images of fossils to build and research image classifier and detection models. This report includes recommendations for data collection, organization, and long-term storage, along with example images and a proposed folder architecture. We also provide a concise overview of the neural network models evaluated during the project.

Table of contents

1 Recommendations for Data Collection	2
1.1 Example Images	2
1.1.1 Good Images	3
1.1.2 Not-so-good Images	3
1.2 Steps for Image Cleaning	4
1.2.1 Example	4
1.3 Creating a Balanced Dataset	4
2 Proposed Data Storage	5
2.1 Folder Structure	5
3 Research Findings	6
3.1 Computer Vision Models	6
3.2 Performance Metrics	6
4 Helpful Links & Resources	7

1 Recommendations for Data Collection

1.1 Example Images

The existing photos received from The Whitney Lab came from three folders: “TBD Microfossils”, “Completed Microfossils”, and “Burke Microfossils”. The “Completed Microfossils” folder had the cleanest images and most consistent format. Figure 1 below shows a glimpse of the variety in resolution, background, aspect ratio, and miscellaneous objects.

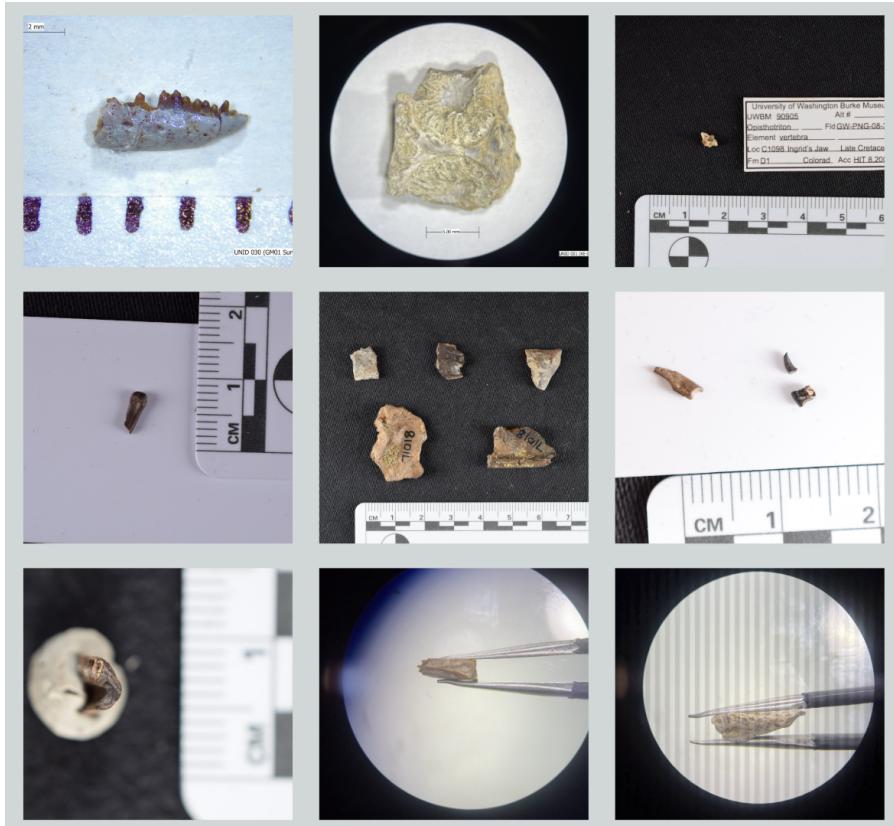


Figure 1: Existing Images

We used the following selection criteria when building our training and testing sets:

- Background: White or black
- Focus: Image could be cropped to contain only the fossil
- Resolution: No strict guideline, as long as features can be seen reasonably
- Distance: No strict guideline, as long as features can be seen reasonably

You can view examples of what we did and did not include in our final dataset in Figures 2 and 3.

1.1.1 Good Images



Figure 2: Images Selected for Modeling

1.1.2 Not-so-good Images



Figure 3: Images not Selected for Modeling

1.2 Steps for Image Cleaning

1. Using the guidelines outlined above, select images.
2. If needed, crop the image to remove miscellaneous objects or increase visibility of features.

1.2.1 Example

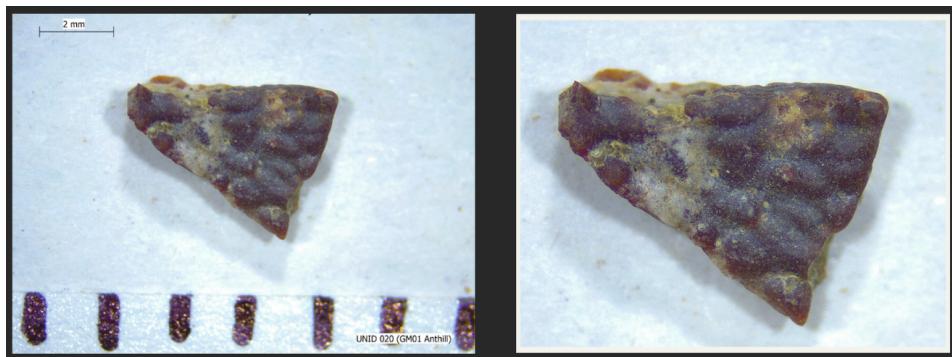


Figure 4: Image Cleaning

1.3 Creating a Balanced Dataset

For image detection and classification models, it is best to have a high number of images with balanced classes (fossils and rocks). This is to try to prevent model bias towards whichever class is found more frequently in the data. More information on why this matters and how to deal with imbalanced data can be found in the **Research Findings** and **Helpful Links & Resources** sections.

The Whitney Lab can help increase the accuracy of models by collecting more photos of rocks and fossils, keeping in mind that we want approximately an equal number in each category. We can increase sample size by using online images, however, we believe that the model will be strongest and most useful if trained primarily on the rocks and fossils collected from dig sites or regions the lab is interested in studying. For a robust model, it's recommended to have at least 1,000 images per class.

2 Proposed Data Storage

2.1 Folder Structure

The following structure is proposed to organize both existing and modeling-ready images for the lab fossils project.

- **raw/** contains all existing images for archival purposes
- **Images_for_Modeling/** contains images prepared for training and classification, separated by class (class1 = fossil, class2 = rock)

Microfossil Curation ▾    

```
raw/
    TBD Microfossils/
    Completed Microfossils/
    Burke Microfossils/

Images_for_Modeling/
    class1/
    class2/
```

3 Research Findings

3.1 Computer Vision Models

We compared three classification models on the provided data. The primary objective of a classification model is to predict the class of a single object given an image of that object.

The long-term goal is to develop a detection model, which can identify and classify multiple objects within a single image. Detection models typically build upon classification models, so it is crucial first to determine which classification model performs best. Selecting the strongest classification model provides a solid foundation for future work in object detection.

Models Tested:

- EfficientNet
- ResNet50
- InceptionV3

3.2 Performance Metrics

EfficientNet

- Highest fossil recall (0.94)
- decent accuracy (0.65)
- This model can identify almost all fossils correctly, but classifies ~50% of rocks as fossils. essentially, at this point, it could help rule out about half of rocks

ResNet50

- lowest fossil recall (0.67)
- poor accuracy (0.30)
- This model is not much better than guessing

InceptionV3

- high fossil recall (0.89)
- highest accuracy (0.93)
- This model has the highest overall accuracy, but misses some fossils.

Takeaways: EfficientNet and InceptionV3 are the most promising model types. After obtaining more data, these models should be explored more thoroughly. ResNet50 can be ruled out as a good model for this project.

4 Helpful Links & Resources

For more on balanced data, and dataset size:

[Roboflow: Training Data Best Practices](#)

For more on the classification models tested:

[EfficientNet](#)

[ResNet50](#)

[InceptionV3](#)