



Experiential Introduction to AI and Ecology

Final Project Proposal

The Beetle Group

Beetle Identification and Trait Segmentation

Team

Alyson East, Sam Stevens, S M Rayeed

Lead: **Professor Sydne Record**

December 13, 2024

1 Background and Relevance

The National Ecological Observatory Network (NEON) field sites collect ground beetle samples annually through a coordinated and standardized sampling effort across the United States. These samples represent a treasure trove of ecological data, yet processing the specimens comes with significant challenges. The most pressing issue is the manual species identification process, which is both time-intensive and tedious, often requiring between 45 minutes and several hours per specimen. This bottleneck not only delays data processing but also limits the potential to extract additional valuable information from these specimens. Beyond identification, there is a wealth of untapped trait data that could provide critical insights into beetle morphology, ecology, and evolution. Traits like body size, elytra length, and thorax dimensions are essential for understanding species' roles within ecosystems and their responses to environmental changes. However, extracting these traits from tens of thousands of archived specimens in the NEON biorepository has been largely infeasible due to the labor required. This has left a substantial gap in our ability to fully leverage the NEON dataset to address broader biodiversity and conservation questions. By harnessing advancements in AI and machine learning, we can automate both species identification and trait measurement, unlocking new possibilities for high-throughput ecological research and enabling insights that were previously unattainable.

Recent advancements in computer vision and foundation models offer transformative tools to address these challenges. For morphological trait measurement, cutting-edge computer vision techniques for edge detection can help accurately delineate the boundaries and key points of beetle body parts, preserving their morphological shapes. By leveraging the Segment Anything Model (SAM), we can further refine this process, seamlessly segmenting beetle features like the elytra, thorax, and abdomen with minimal hyperparameter tuning. These segmentations enable precise trait measurements, such as elytra length and width, which are crucial for understanding ecological and evolutionary variations. For species identification, foundation models like BioCLIP and BLIP present promising solutions. Models like BioCLIP, being trained on diverse datasets like TreeOfLife-10M, excel in zero-shot and open-ended classification tasks, making them especially valuable for identifying Hawaiian-specific beetles such as Ambrosia and Rhinoceros beetles. With limited labeled data available for these species, fine-tuning these foundation models can yield robust predictions, accelerating species identification workflows. Integrating these AI tools can streamline

both morphological analysis and species identification, unlocking new possibilities for ecological research across NEON datasets and beyond.

2 Research Questions

Our core research question is: how can AI-based tools streamline the study of Hawaiian beetle species? Specifically, we attack two separate problems:

1. Morphological trait segmentation
2. Invasive species classification

Both problems have deep biological motivation, and answering questions about these topics is limited by the speed at which we can collect relevant data. They are both problems for which AI-based automation would be broadly useful and offer a useful testbed to develop and analyze better tools. We describe the two problems in more detail below.

2.1 Morphological Trait Segmentation

We focus on this morphological trait segmentation because morphological traits play a fundamental role in shaping ecological communities. Traits, such as body size, directly influence how species interact with their environment and with one another. They determine which species can persist under specific environmental conditions (external filters, like temperature or habitat) and how they coexist or compete within a community (internal filters, such as predation or competition). However, manually measuring these traits, especially for large datasets such as the beetles, is incredibly time-intensive and limits our ability to fully explore their ecological significance. By automating this process, we can harness trait data more effectively to understand how species traits shape ecological diversity, structure communities, and guide conservation efforts.

2.2 Invasive Species Classification

Invasive Hawaiian beetle species, such as ambrosia and rhinoceros beetles, have profound ecological and socioeconomic impacts. Ambrosia beetles, for example, are the primary drivers of Rapid 'Ōhi'a Death (ROD), a devastating disease affecting the 'ōhi'a *lehua*, a keystone tree species critical to Hawaiian ecosystems. With six ambrosia beetle species predominantly causing ROD, their identification and management are crucial for protecting native forests and the biodiversity they support. Similarly, rhinoceros beetles are notorious for causing significant damage to coconut palms and other agricultural crops, threatening not only native ecosystems but also Hawaii's agricultural economy and cultural heritage. Their broader effects resonate through both biological



Figure 1: From left: (1) Sample ambrosia beetle ¹, (2) Rapid ‘Ōhi‘a Death caused by ambrosia beetle ², (3) Sample coconut rhinoceros beetle ³, (4) palm damage caused by CRB ⁴

and social systems, making accurate identification and monitoring critically important. By fine-tuning foundation models like BioCLIP, and working closely with the Intake project group, we aim to develop efficient and scalable tools for tackling the invasive Hawaiian beetle species identification challenge.

3 Project Goals

Morphological Trait Segmentation Develop a pipeline for automated segmentation and precise measurement of key morphological traits in Hawaiian beetles, such as body size and elytra length. By using advanced computer vision techniques and the Segment Anything Model (SAM), we aim to enhance trait data extraction from large datasets, enabling rapid analysis of ecological diversity and species roles.

AI-Driven Beetle Identification Fine-tune foundation models like BioCLIP for the identification of invasive Hawaiian beetles, including Ambrosia and Rhinoceros beetles. These species pose significant ecological and economic threats, and by improving model accuracy with domain-specific data, we aim to streamline identification and support early detection efforts.

Scalability and Integration Create a scalable framework for processing and analyzing large volumes of archived specimens from the NEON biorepository. This framework will allow for seamless integration of future specimens, ensuring consistent and real-time data processing, and will be adaptable to other species and regions for broader ecological research.

4 Methods

We primarily will be imaging existing beetle specimens, annotating data, and developing AI-based tooling to automatically annotate future data.

4.1 Fieldwork

Given that beetles are not sampled by NEON in the winter at the Pu'u Maka'ala Natural Area Reserve (PUUM) site, we will not be collecting new pit trap samples during our time in Hawaii, but instead working with samples that have already been collected and processed by the NEON field team from the field collection season.

4.2 Lab work

Existing beetle specimens will need to be imaged, digitized, and linked to metadata. Photos of the beetles will be taken in batches with a DSLR camera equipped with a macro lens for imaging small specimens ($>4\text{mm}$). Specimens will then need to be processed to generate files with photos of individual beetles and linked to the appropriate metadata (Genus, Species, Sample Location, etc.). A stratified random subset of the beetle images will then need to be manually annotated for traits of interest to train the segmentation model. We plan on annotating the following traits:

- Full body size
- Abdomen size
- Thorax size
- Elytra length and width

4.3 AI Methods

Foundation models are the premier method for developing new AI-based tools; foundation model generality means they can be applied to nearly any task without expensive training. We will primarily leverage three existing foundation models, each for different tasks.

Segment Anything Model (SAM) for morphological Trait Segmentation SAM is a foundation model developed by Meta that separates images into “parts”. We will use SAM to automatically delineate key beetle features like elytra, thorax, and abdomen. We can then precisely measure traits, including body size and other morphological attributes, which are essential for ecological analysis.

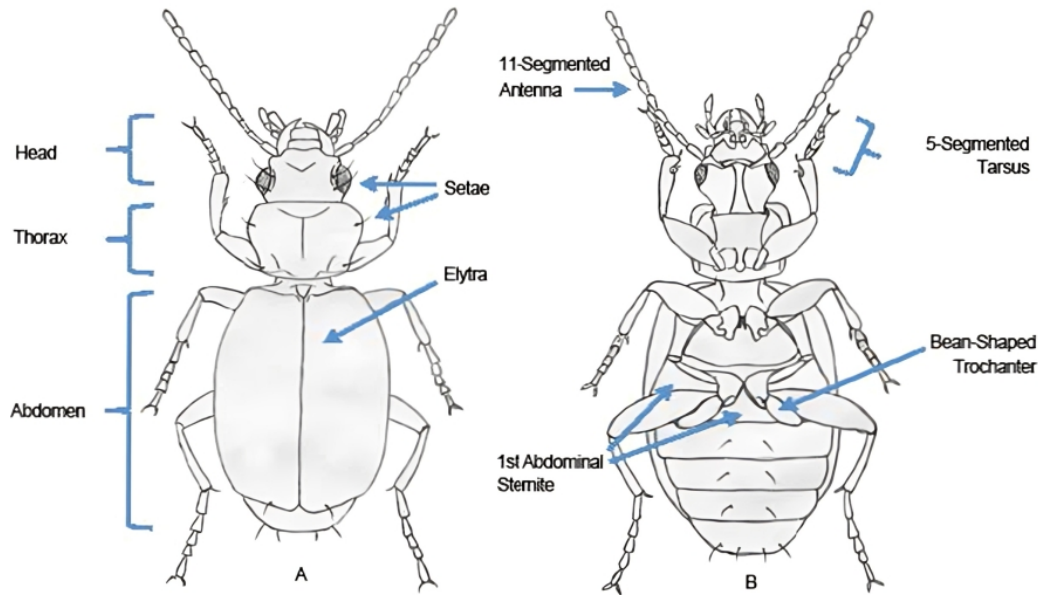


Figure 2: Example of ground beetle body traits (A: Dorsal, B: Ventral) ^{source}

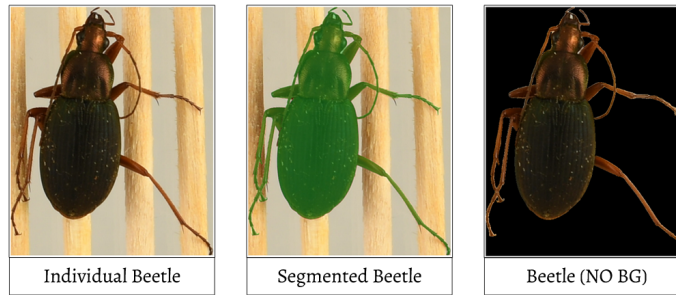
BioCLIP for Species Recognition BioCLIP is a foundation model developed by Imageomics for species classification of living organisms. We will leverage BioCLIP to automatically identify Hawaiian beetle species. We will fine-tune BioCLIP with task-specific data to develop a strong beetle classifier for invasive species like Ambrosia and Rhinoceros beetles.

YOLO for Species Detection The YOLO series of models are general object detection models. We will use YOLO to detect beetles and place bounding boxes around them in images. This more precise localization will facilitate more focused and accurate species identification, especially in cluttered or complex environments.

5 Relevant Prior Work: Easy Traits - BeetlePalooza'24 (Link)

The Easy Traits project, showcased at BeetlePalooza'24, aimed to develop a pipeline for the automated morphological trait analysis of ground beetles. While we focused on a very small subset of image samples, it demonstrated the core capabilities of our proposed pipeline, which can be helpful in our current work:

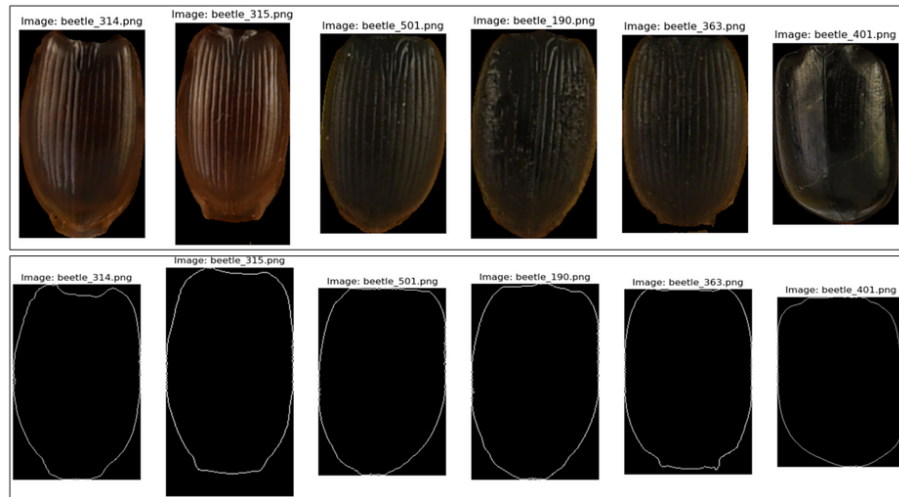
1. **Segmenting Individual Beetles:** Using SAM, we successfully isolated individual beetles from specimen images containing many beetles. It was crucial to reduce noise and ensure accurate subsequent analysis. YOLO could also achieve this goal.
2. **Background Removal:** After segmenting the beetles, the background was removed (replaced with a black background) to create clean, standalone images. This eliminated extraneous visual elements and focused on the beetle's anatomical features.



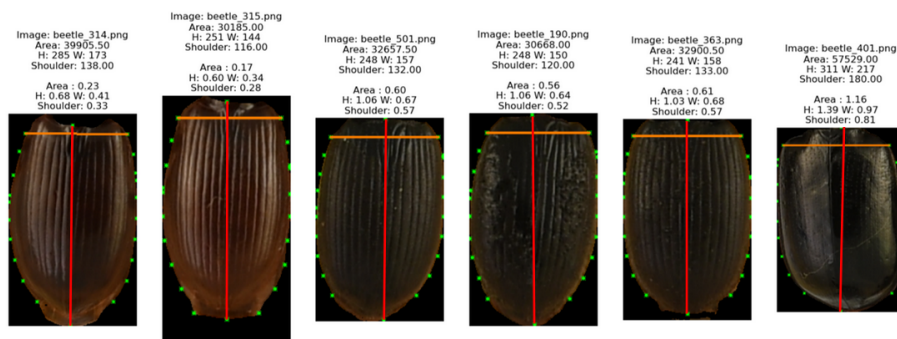
3. **Elytra Segmentation:** A second-tier segmentation on the elytra was performed. This required fine-tuning the SAM model to recognize and delineate these specific structures accurately (We implemented it by extracting the area with maximum size out of all - in almost all cases, it was the elytra).



4. **Edge Detection of Elytra:** To facilitate quantitative measurements, the edges or boundaries of the segmented elytra were identified. By employing several robust edge detection algorithms (and comparing the performance), the pipeline ensured high fidelity in defining the elytra's contours.
5. **Landmark Identification and Trait Measurement:** Key landmark points were determined across partitions of the elytra boundary, including leftmost, rightmost, topmost, and bottom-most points in specified segments. Using these landmarks, the pipeline measured two critical traits: elytra length and width.



6. **(Derived) Trait Measurement:** To convert pixel values into real-world units, we used a ruler/scale/checker-box for calibration. Pixel values can vary based on image scale—close-up images capture more pixels, while distant shots result in fewer. Calibration ensures accurate and consistent trait measurements like elytra length and width, regardless of image distance.



6 Relevance to the Course Theme

This project sits at the intersection of computer science, machine learning, and ecology, embodying the potential of AI and computational tools to tackle complex environmental challenges. By automating the identification of beetle species and the measurement of morphological traits, we are leveraging cutting-edge computer vision techniques to make significant strides in biodiversity research. Applying models like BioCLIP and the SAM to ecological data reflects how AI can

transform traditional ecological research methods, enabling more efficient data processing, better scalability, and faster analysis across large datasets. This project aligns with the course's emphasis on utilizing AI and machine learning to address real-world environmental issues. The ability to quickly and accurately identify invasive beetle species (ambrosia and rhinoceros) has profound ecological and socioeconomic implications, as they are responsible for the devastation of native ecosystems in Hawaii, highlighting the importance of automating detection and monitoring processes. By incorporating AI with ecological research, this work opens up new avenues for large-scale biodiversity monitoring and conservation efforts, facilitating a deeper understanding of how species interact within ecosystems, evolve, and adapt to environmental changes. Furthermore, the project's focus on fine-tuning machine learning models for specific ecological challenges underscores the role of interdisciplinary approaches in advancing both computational science and ecological sustainability.

7 Expected Results

The primary outcome of this project will be a software pipeline capable of automated morphological trait segmentation, preferably focusing on Hawaiian beetles. The segmentation model SAM is expected to provide high-precision large-scale measurement of traits, which will offer critical insights into beetle morphology, ecological roles, and their potential responses to environmental shifts. In parallel, the invasive species identification (and monitoring) is of significant concern due to their ecological and economic impacts. Beyond species identification, the AI tools developed in this project will lay the groundwork for future large-scale applications, as the pipeline will be scalable to process specimens from the entire NEON biorepository, potentially extending to other ecosystems and taxa. These outcomes will provide a comprehensive solution to the bottleneck of manual trait measurement and species identification, paving the way for efficient, data-driven ecological research. Moreover, the project's results are expected to make a significant contribution to the broader goals of biodiversity conservation and ecosystem management by offering tools for monitoring species distribution, assessing ecological health, and understanding species' evolutionary adaptations. We hope this project will enhance our ability to conduct high-throughput biodiversity studies and make informed decisions in conservation practices, addressing both scientific and practical needs within the field of ecology.