Tracking Aquatic Insect Emergence Using Machine Learning

From Manual classifying work to a complete automated solution that includes a ready-to-deploy new website linked with our pipeline.

Project Lead: Dr.Emily Bernhardt, **Project Manager:** Dr.Heili Lowman

Team members: Abdulmlik almuhanna, Boyu tan, melosa rao, Uzair Chaudary

Table of Contents

Tracking Aquatic Insect Emergence Using Machine Learning..... 1
Project Objectives and Goals..... 3
Current Progress and Milestones Achieved..... 3
Preliminary Results..... 4
Challenges and Obstacles Encountered..... 4
Next Steps and Future Work..... 4
Resources and Tools Used..... 5

Project Background

Aquatic insects are crucial to stream ecosystems, acting as decomposers and a vital food source for various organisms. Their emergence patterns reflect ecosystem health and help scientists understand environmental changes. Traditionally, Tracking their numbers has been done by collecting and identifying these insects which is a labor-intensive process that involves trapping, drying, packaging, shipping, and manual classification.

Project Objectives and Goals

Automate the end-to-end identification of aquatic insects from images of sticky traps to replace a labor-intensive manual process that previously required ~10 days per trap and ~600 hours per year in the Hubbard Brook experiment. The targets are to (i) achieve high-recall detection of very small insects (<1 mm), (ii) deliver reliable multi-class classification despite heavy class imbalance, (iii) distinguish insects from non-insect debris, and (iv) reduce annual effort to ~20 hours while enabling rapid, scalable ecological monitoring.

Current Progress and Milestones Achieved

A two-phase pipeline has been implemented. In Phase 1 (detection), a YOLO variant tailored for small objects (YOLO-P2) combined with SAHI algorithm that provides a high-resolution Window-Sliding strategy, increased recall from 46% to approximately 87%. Negative images and extensive hyperparameter tuning further improved robustness. In Phase 2 (classification), a ResNet-based model—trained with targeted augmentations and under-sampling of dominant classes—reached ~96% accuracy across most classes, with remaining gaps in the heterogeneous “Other” class. A Flask web interface was built to expose the pipeline to end users, integrated with Roboflow for seamless upload, inference, annotation storage, and rapid dataset growth. The team finalized a professional poster and presented outcomes in stakeholder and faculty sessions.

Preliminary Results

Detection achieved ~87% recall on validation data after adopting the window-sliding strategy and YOLO-P2. Classification attained ~96% accuracy overall, though performance on the broad "Other" category remains comparatively weak due to intra-class heterogeneity. Qualitative reviews indicate substantial reduction in false negatives for tiny insects and practical end-to-end usability through the web interface and automated dataset curation.

Challenges and Obstacles Encountered

Key challenges include (i) detecting sub-millimeter insects against noisy backgrounds, (ii) severe class imbalance with some taxa observed only a few dozen times versus thousands for others, (iii) distinguishing insects from visually similar debris such as dirt or leaf fragments, and (iv) the catch-all "Other" class, whose internal diversity complicates learning a stable decision boundary.

Next Steps and Future Work

Planned improvements focus on (i) restructuring the "Other" category into finer-grained labels or hierarchical classification, (ii) experimenting with alternative backbones (e.g., transformer-based detectors/classifiers) to further boost small-object sensitivity, (iii) targeted data collection and augmentation for under-represented taxa, (iv) active-learning loops to prioritize uncertain cases for human review, and (v) continued productization of the web app (logging, monitoring, and deployment hardening).

Resources and Tools Used

- http://hbwater.org:3838/data_stories/emergence/
- https://bigdata.duke.edu/projects/tracking-aquatic-insect-emergence-using-machine-learning-climate/
- https://arxiv.org/abs/2104.10419
- https://arxiv.org/abs/1512.03385
- https://www.researchgate.net/publication/358603657_Slicing_Aided_Hyper_Inference_and_Fine-tuning_for_Small_Object_Detection
- https://flask.palletsprojects.com/en/stable/