Machine Learning for Ecology: Automated Invertebrate Monitoring on Large Rivers

Liam Zarri[1,2], Nick Macias[1,2], Eric Orenstein[3], Eric Danner[2], Eric Palkovacs[1]
1: UC Santa Cruz, 2: National Marine Fisheries Service, 3: Scripps Institute of Oceanography

Background

- Insect drift is a fundamental mechanism within freshwater ecosystems for the distribution of genetic material, predator avoidance, and the downstream colonization of open niche space (Fonseca 1999)
- Endangered fish species on the Sacramento River, such as chinook & green sturgeon, depend on drift during early life stages (Dodrill et al. 2016)
- Characterizing the timing and abundance of drift can assist in management of Endangered Species Act (ESA) listed species
- Current drift sampling is only practical in wadeable waters, not in the thalweg (middle of the river channel)
- Current drift net sample processing is labor-intensive

Objective

- In collaboration with the Jaffe Lab at Scripps, develop camera technology to sample drift invertebrates in the thalweg and capture temporal fluctuations
- Develop supervised learning machine learning software to classify images and generate data on taxa , size, and drift timing

Dodrill, M.J., Yackulic, C.B., Kennedy, T.A., Hayes, J.W. 2016. Prey size and availability limits maximum size of rainbow trout in a large tailwater: insights from a drift-foraging bioenergetics model. Can. J. Fish. Aquat. Sci., 73: 759-772.
Fonseca, D.M. 1999. Fluid-mediated dispersal in streams: models of settlement from the drift. Oecologia, 121: 212-223.

Methods

Figure 1. Camera deployed in the Sacramento River

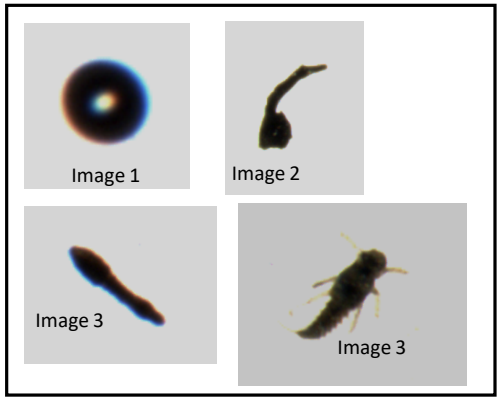

Figure 2. Raw images from the camera

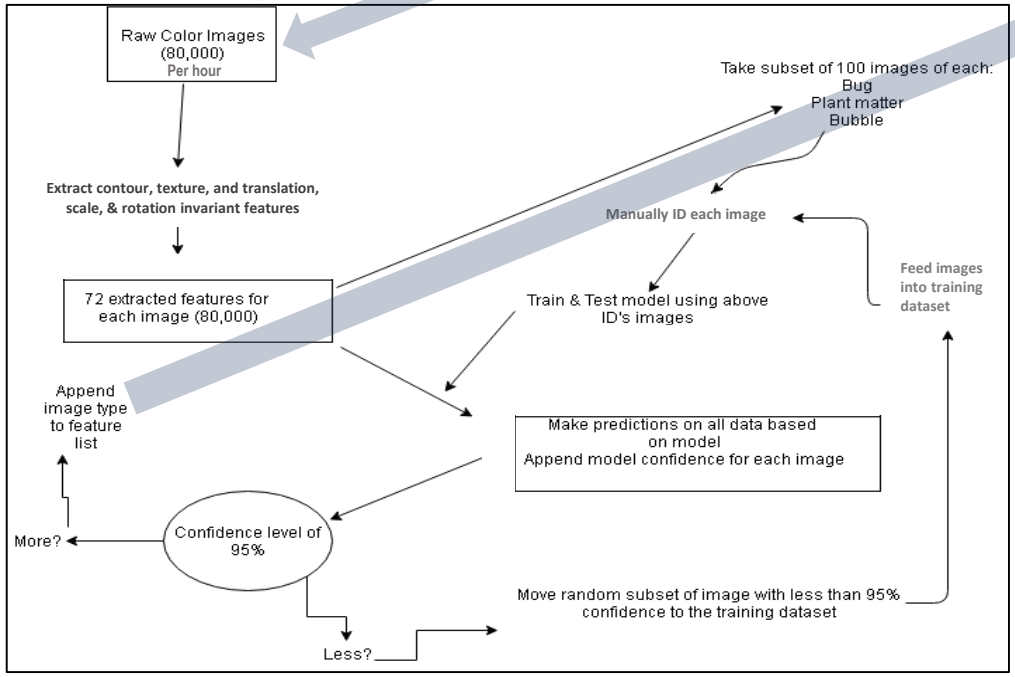

Figure 3. Flowchart of machine learning script, developed in python 2.7 with OpenCV 2

Results

- Images are binned into 3 classes: invertebrate, bubble, or plant matter
- Script identifies 85% of images with >95% accuracy; only 0.2% of images are invertebrates
- This will increase as training dataset of invertebrates increases; at this point there are only 734 invertebrate images

ID	Time of Capture	Length (mm)	Image ID	Confidence
1	12:01 5/7/2017	0.4	Bubble	97%
2	12:01 5/7/2017	2.3	Plant Matter	95%
3	12:04 5/7/2017	3.3	Invertebrate	99%
4	12:04 5/7/2017	2.5	Invertebrate	98%

Figure 4. Sample of data output from machine learning script

Conclusions and future directions

Applications of this tool include:

- capability to sample at greater depth & flow than drift nets, which expands sampling efforts into previously unsampled high velocity thalweg
- Continuous sampling correlated with onboard environmental sensors (flow, temperature, ect.)
- Autonomous deployment to provide continuous data across broad temporal scales
- Significant reduction of technician time & cost required to process samples
- Sampling in turbid or nighttime conditions

This tool could assist in quantifying the response of drift invertebrates to management actions such as dam releases, thus impacting ESA-listed species in large river systems.