

Observing Siphonophores: Accelerating Behavioral Observations with Machine Learning

Kira Kaplan, Smith College

Mentors: Kakani Katija, Joost Daniels, Eric Orenstein, Paul Roberts

Summer 2023

Keywords: siphonophore, behavioral observation, machine learning

ABSTRACT

Documenting the behavior of siphonophores presents a considerable challenge without any clear methodology. Observations rely primarily on in situ ROV video data since siphonophores have proved difficult to work with in an ex situ environment. The Monterey Bay Aquarium Research Institute (MBARI) has over 30 years of ROV footage that can be used to make these observations. However, the process for using this footage to make behavioral observations is extremely time and labor intensive. The *anno2trackstart* pipeline was created to reduce this time constraint and streamline the process. This pipeline utilizes a machine learning model to accurately predict the initial sighting of a siphonophore within a video clip. This paper describes the pipeline process as well as the accuracy of a machine learning model in detecting and classifying siphonophores. Findings suggest that machine learning models and the *anno2trackstart* pipeline can be a successful alternative way of documenting siphonophore behavior.

INTRODUCTION

Siphonophores are gelatinous, colonial animals under the phylum Cnidaria that live in the midwater region of the ocean (Mapstone 2014). Despite being a central predator in the

midwater food web (Choy et al. 2017), very little is known about their behavior and functionality. Their soft-bodied composition results in extreme fragility and presents significant challenges to collection for ex situ use. Additionally, after collection they do poorly in captivity and exhibit modified behavior due to stressful conditions. For these reasons, the majority of siphonophore research relies on in situ observations.

MBARI has over 30 years of labeled in situ ROV and AUV transect video data stored in MBARI's Video Annotation and Reference System (VARS). Each organism that appears in a video is manually logged in VARS with identifying and contextual information from the ROV such as depth, location, temperature, etc. (Robison et al. 2017). Each annotation also contains a timestamp with the date and time of the sighting. While these annotation timestamps are key in making behavioral observations, there exists one caveat: the timestamps do not occur at the initial sighting of the organism.

When making behavioral observations based on transect video data, the observation should be made based on the initial sighting of the siphonophore. As the ROV moves closer to the organism, there is a greater chance that vehicle disturbance will disrupt its natural behavior. Some siphonophores have shown to be sensitive to light (Sutherland et al. 2019) and changes in current. Therefore, to ensure the most accurate representation of natural siphonophore behavior, the observation should be made at its initial sighting to avoid potential disruption.

Another challenge in making these observations lies in the tedious and convoluted process of using an annotation in VARS to find the specific observation period. This paper will discuss a simplified approach to making these observations using machine learning.

MATERIALS AND METHODS

MACHINE LEARNING MODEL

To identify the initial sighting of a siphonophore from a video clip, a YOLOv5 machine learning model pre-trained on MBARI's 315K dataset (a dataset comprised of 315K images of animals in the ocean) was used. This allowed the model to identify and classify

a large variety of different organisms, including siphonophores. This pre-trained model was used to create object tracks for siphonophores. These object tracks (Figure 1) contain information about the siphonophore from the first object detection on screen up until the last. Using the outputted object tracks, the first detection of the organism (otherwise known as the t_0) can be used to make the behavioral observation.

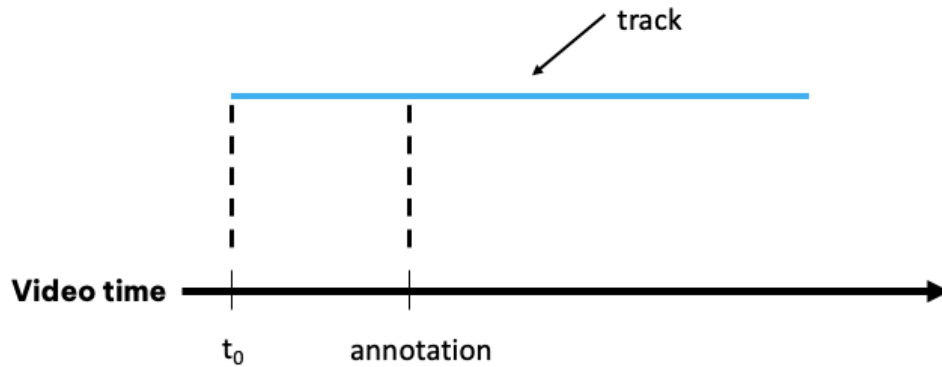


Figure 1. Annotation and t_0 timestamp for an object track generated by a model for a video clip.

A video clip serves as the input for the model, and the output includes (1) the same video clip with bounding boxes and (2) a tracking text file. The text file lists the object detections and their classifications for each frame in the video. This file is used to find the t_0 based on the first detection of a siphonophore.

PIPELINE

To expedite and streamline the process of making these behavioral observations, the Python *anno2trackstart* pipeline (source code for *anno2trackstart* is available at <https://github.com/bioinspirlab/anno2trackstart>) was written. A simplified process diagram for this pipeline can be seen in Figure 2. Before running *anno2trackstart*, a dataset is formed by querying the VARS database for annotations of unique siphonophores. These annotations are collected and stored as a csv file. The first half of the *anno2trackstart* pipeline takes the csv file as its argument and iterates through the annotations in the csv file. For each annotation, the path of the original video containing

the annotation is parsed and added as a new variable to the csv file. The path and annotation timestamp are then used to trim the original full-length video (typically around 15 minutes long) to include 10 seconds before the annotation timestamp and 10 seconds after the annotation timestamp, for a total clip of 20 seconds. This new, shortened video decreases the processing time it takes for the model to create an object track and limits the number of additional organisms within the video. The trimmed clip is saved in an “Untracked” folder in the working directory and then run through the model with a confidence threshold of 0.25. The confidence threshold takes on a value between 0-1 and indicates how confident the model must be to detect and classify an object. Therefore, the model will only track objects when it has 25% or more confidence in its detection. The model outputs a video with bounding boxes and a tracking text file that are saved in a “Tracked” folder. Once all the annotations have been found, trimmed, and tracked, the second half of the pipeline is run.

The second half of the *anno2trackstart* pipeline takes the path to the “Tracked” folder (containing the object tracks for each annotation) as an argument. For each annotation, it searches through the tracking text file and finds the first classification of ‘Siphonophorae.’ The video frame associated with this detection is the t_0 . The t_0 frame number is turned into a timestamp and saved in the csv file as a new variable. A direct link to this timestamp in the original video file is created and saved in the csv file for easy access. Once the pipeline finishes running, observations may be manually recorded using the direct links available in the csv file.

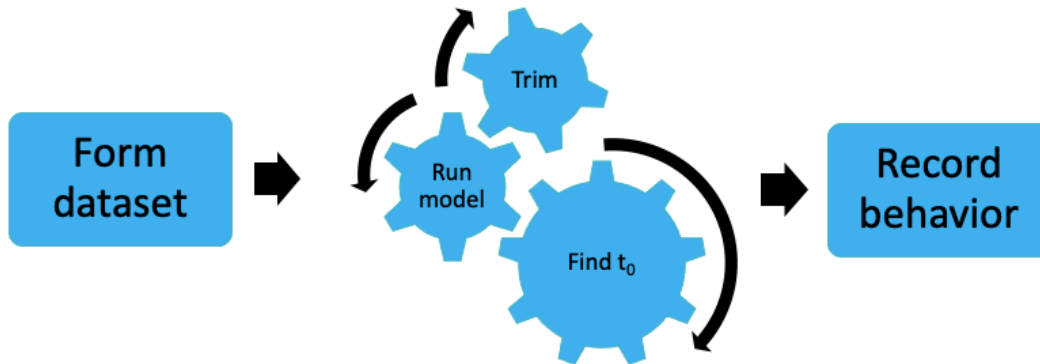


Figure 2. Simplified process diagram for recording behavioral observations with the *anno2trackstart* pipeline.

TESTING FOR ACCURACY

To ensure the model’s ability to detect siphonophores and correctly identify the t_0 , a series of accuracy tests were conducted. A collection of 42 distinct annotations, split evenly among three siphonophores—*Praya dubia*, *Erenna*, and *Apolemia*—were queried from the VARS database (Figure 3).



Figure 3. *Praya dubia* (left), *Apolemia* (middle), and *Erenna* (right.)

For each annotation, the model’s t_0 was compared to the true t_0 . The true t_0 was obtained by manually finding the first frame at which the human eye could identify and classify the object onscreen as a siphonophore. The difference in the model and true t_0 were analyzed and compared between the three siphonophore groups using R. Additionally, two different confidence thresholds (0.25 and 0.55) were compared using the same process with 14 annotations of *Praya dubia*.

RESULTS

A comparison between the model's ability to predict t_0 at a 0.25 and a 0.55 confidence threshold is shown in Table 1.

Confidence Threshold	Δt_0	Standard Deviation
0.55	39.36 frames or 1.31 seconds	24.20 frames or 0.81 seconds
0.25	13.36 frames or 0.45 seconds	21.32 frames or 0.71 seconds

Table 1. The average difference and standard deviation between the model's t_0 and the true t_0 for 14 *Praya dubia* annotations run through with 0.25 and 0.55 confidence thresholds.

The difference is statistically significant ($t(25.59) = -3.02$, $p = 0.0057$) and appears visually different when observed. Figure 4 displays a side-by-side comparison of a model t_0 for the same annotation run through both confidence thresholds. It's visually apparent that the model run at a 0.25 confidence threshold can detect the object as a siphonophore much sooner and also recognize more of the organism. This supports the conclusion that the 0.25 confidence threshold should be used over 0.55.



Figure 4. The video frame at t_0 for a video clip run with 0.25 confidence threshold (left) and 0.55 confidence threshold (right).

When looking at the model’s ability (run at 0.25 confidence) to predict an accurate t_0 , success varied amongst siphonophore species as shown in Table 2. The model performed much better against Praya dubia and Apolemia annotations than Erenna annotations.

Siphonophore	Δt_0	Standard Deviation
Praya dubia	13.36 frames or 0.45 seconds	21.32 frames or 0.71 seconds
Apolemia	21.15 frames or 0.71 seconds	42.70 frames or 1.42 seconds
Erenna	119.5 frames or 3.99 seconds	133.94 frames or 4.47 seconds

Table 2. The average difference and standard deviation between the model’s t_0 and the true t_0 for 14 Praya dubia, Apolemia, and Erenna annotations.

DISCUSSION

When looking at the differences in model accuracy between the siphonophore groups, there are several factors that could be attributed. Within the annotations chosen for the test analysis, the majority of Erenna samples were taken from older videos with reduced video quality. This is due to limited availability of Erenna annotations within the VARS database (Figure 5).

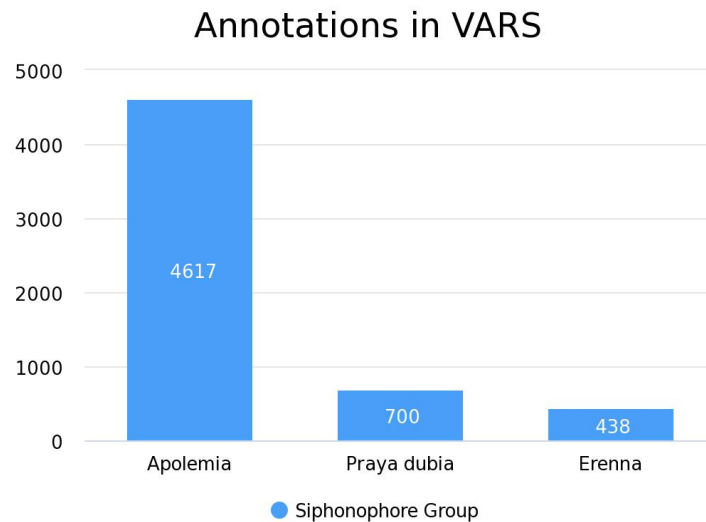


Figure 5. The number of unique annotations for each siphonophore group in VARS.

Many of the Erenna annotations available in the database are linked to the same organism as ROV's tend to linger and spend a longer time obtaining footage of Erenna due to its rarity. This further reduces the number of annotations available for use of unique individuals. This restricted Erenna dataset suggests that within the 315K dataset (the training set of images used to train the YOLOv5 model) there are fewer Erenna images. This presents a challenge for the model (trained with insufficient Erenna data) to identify these organisms. Additionally, the poorer quality in videos used for this sample set may contribute to and decreased and biased accuracy.

Despite these differences in accuracy, there does not seem to be a significant difference between the model's t_0 and the true t_0 . Additionally, the model's t_0 occurs at a point in the video where the organism is far enough away from the ROV to still avoid vehicle disturbance. The accuracy variability does not indicate that the model approach should only be used with species at which it performs perfectly, but rather that additional features might be put in place when working with different siphonophores. For example, when working with annotations from Erenna, adding an extra 3 seconds to the model's t_0 can ensure the initial sighting is captured. This 'time buffer' can be altered within the pipeline to better suit a variety of species.

It is worth mentioning that the three siphonophores used in this experiment share morphological similarities. Both Erenna and Apolemia belong to the siphonophore suborder Physonectae, and Praya dubia to the Calycophorae suborder. The accuracy of the model and its success in the *anno2trackstart* pipeline therefore may vary with morphologically different species.

CONCLUSIONS/RECOMMENDATIONS

The pipeline used in this project is a viable alternative to the manual process of recording behavioral observations for siphonophores. It greatly reduces the amount of labor required, but whether it shortens the time remains to be seen. A time comparison should be run by completing the manual steps to record behavioral observations for a set of annotations, and then running that same set of annotations through the *anno2trackstart*

pipeline. Once a comparison is made, a more concrete suggestion can be made for whether this automated process should be used as a replacement. Regardless, the machine learning model is sufficient in locating the initial sighting of a siphonophore in a video and can be used as a means of finding the t_0 . To improve the accuracy of this prediction, a model can be trained on an image set with a greater number of siphonophore images, including a wider diversity of species.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my mentors Drs. Kakani Katija, Joost Daniels, Eric Orenstein and Paul Roberts for all their guidance and support in making this project possible. Much gratitude to the Video Lab for all their work in annotating and labeling the ROV and AUV data, without which this project would not exist. I'd also like to thank Lonny Lundsten, Brian Schlining, and Kevin Barnard for their support in navigating VARS, the YOLOv5 machine learning model, and countless other MBARI software applications. I'd especially like to thank Dr. George Matsumoto for organizing the 2023 MBARI summer internship program and making this opportunity possible. Lastly, I would like to thank NSF-IOS #2114169 (to Kakani Katija), the David and Lucile Packard Foundation, the Dean and Helen Witter Family Fund, the Rentschler Family Fund, the Maxwell/Hanrahan Foundation, and the Monterey Bay Aquarium Research Institute for providing the resources to support this project.

References:

- Barham, E. G. (1963). Siphonophores and the Deep Scattering Layer. *Science*, 140(3568), 826–828. <http://www.jstor.org/stable/1711316>
- Choy, C.A., S.H.D. Haddock, and B.H. Robison. 2017. Deep pelagic food web structure as revealed by in situ feeding observations. *Proc Biol Sci*, 284: 1–10. <http://doi.org/10.1098/rspb.2017.2116>
- Costello, J., Colin, S., Gemmell, B. *et al.* Multi-jet propulsion organized by clonal development in a colonial siphonophore. *Nat Commun* 6(8158). <https://doi.org/10.1038/ncomms9158>
- Kelly R. Sutherland, Brad J. Gemmell, Sean P. Colin, John H. Costello; Propulsive design principles in a multi-jet siphonophore. *J Exp Biol* 222(6): jeb198242. <https://doi.org/10.1242/jeb.198242>
- Mackie George Owen. 1964. Analysis of locomotion in a siphonophore colony. *Proc. R. Soc. Lond. B.* 159(975). <https://doi.org/10.1098/rspb.1964.0008>
- Mapstone G. M. (2014). Global diversity and review of Siphonophorae (Cnidaria: Hydrozoa). *PloS one*, 9(2), e87737. <https://doi.org/10.1371/journal.pone.0087737>
- Robison, B.H., K.R. Reisenbichler, and R.E. Sherlock. 2017. The coevolution of midwater research and ROV technology at MBARI. *Oceanography* 30(4):26–37, <https://doi.org/10.5670/oceanog.2017.421>
- Robinson, B.H., K.R. Reisenbichler, R.E. Sherlock, J.M.B. Silguero, and F.P. Chavez. 1998. Seasonal abundance of the siphonophore, *Nanomia bijuga*, in Monterey Bay. *Deep Sea Research Part II: Topical Studies in Oceanography*, 45: 1741-1751. [https://doi.org/10.1016/S0967-0645\(98\)80015-5](https://doi.org/10.1016/S0967-0645(98)80015-5)
- Schlining, B., and N. Jacobsen Stout. 2006. MBARI's Video Annotation and Reference System. *Proceedings of the Marine Technology Society / Institute of Electrical and Electronics Engineers Oceans Conference*.
- Sutherland, K. R., Gemmell, B. J., Colin, S. P., & Costello, J. H. (2019). Maneuvering Performance in the Colonial Siphonophore, *Nanomia bijuga*. *Biomimetics*, 4(3), 62. <https://doi.org/10.3390/biomimetics4030062>