



# **Automated Detection of Deep-Sea Animals**

**Dallas Hollis, California State University Sacramento**

***Mentors: Duane Edgington, Danelle Cline***

***Summer 2016***

**Keywords: Automated Video Event Detection (AVED), Quantitative Video Transects, Video Annotation and Reference System (VARS), NGSS, Education**

## **ABSTRACT**

The Monterey Bay Aquarium Research Institute routinely deploys remotely operated underwater vehicles equipped with high definition cameras for use in scientific studies. Utilizing a video collection of over 22,000 hours and the Video Annotation and Reference System, we have set out to automate the detection and classification of deep-sea animals. This paper serves to explore the pitfalls of automation and suggest possible solutions to automated detection in diverse ecosystems with varying field conditions. Detection was tested using a saliency-based neuromorphic selective attention algorithm. The animals that were not detected were then used to tune saliency parameters. Once objects are detected, cropped images of the animals are then used to build a training set for future automation. Moving forward, neural networks or other methods could be used for automated taxonomic identification of deep-sea animals. With access to greater computing power and a second layer of classification, this method of detection may prove to be very effective in the future for analyzing and classifying large video datasets like the one found at MBARI. Finally, the entire process was used to develop a high school lesson plan that satisfies the Next Generation Science Standards.

## INTRODUCTION

For over 25 years, the Monterey Bay Aquarium Research Institute (MBARI) has deployed remotely operated underwater vehicles (ROVs) approximately 300 times a year.

These video-equipped ROVs have amassed an archive of more than 22,000 hours of video footage. Each dive has been carefully annotated by MBARI professionals with 5.13 million observations using the Video Annotation and Reference System (VARs), an MBARI developed open-source software-hardware. Users can access these annotations by customizing a search query to return detailed information such as temperature, salinity, pH, taxonomy, transect type, and more (Schlining & Stout, 2006). As the collection of videos and associated data has grown, the concept of efficient automated classification has become more appealing (L. Kuhn, personal communication, July 15, 2016).

MBARI software engineers have successfully experimented with the use of Automated Video Event Detection (AVED) in Quantitative Video Transects (QVTs) utilizing a saliency-based neuromorphic selective attention algorithm to detect organisms within a video frame. In winner-takes-all fashion, the most salient objects are segmented and the location of the object in the next frame is then estimated to initiate tracking. Once an object has registered as having a high enough saliency and has been tracked through multiple frames, it is marked as “interesting” and a cropped image of the segmented detection is produced (Walther, Edgington, & Koch, 2004). Building off the success of these two programs and pulling from species collection experience in the field, we wanted to explore the pitfalls of current automation procedures and explore the process of automated detection as it applies to animals in diverse ecosystems with varying field conditions. Finally, the entire process was used as inspiration to create a high school lesson plan that satisfies the current standards as outlined by the Next Generation Science Standards (NGSS).

## MATERIALS AND METHODS

### Video Processing of QVTs

In the past, researchers have used tow nets to conduct species identification and distribution studies. While practical for some applications, the tow net tends to damage gelatinous animals and disrupts natural aggregation patterns. The introduction of QVTs gave researchers the ability to observe these underwater creatures in their natural habitats (Edgington, Cline, Davis, Kerkez, & Mariette, 2006). With the success of the QVT program, the increase in data, and the ability of MBARI to analyze video data from multiple video and image platforms, human analysis quickly became a processing bottleneck. To alleviate the effects of this bottleneck, the following sources of error had to first be identified before procedural steps could be implemented to standardize a processing method.

### Sources of Error in Collection

There were three main sources of error found in prior transcoded video files that had to be corrected before moving forward. Table 1 outlines these errors and briefly describes how they were overcome. First, MBARI dive footage is recorded onto digital HD tapes that are typically changed every hour during an average 6-hour dive. Therefore, each dive will have approximately six digital videotapes associated with it. Typically, before a dive begins, the recording VCR is “zeroed” to read a tape time code of 00:00:00:00 in the format of Hours:Minutes:Seconds:Frames. The first tape for a dive will be set at zero and the second tape will begin where the last tape left off. This creates videotapes that do not start at zero, except for the first video of any particular dive. It was found that a large number of starting tapes are also non-zero due to the VCR not being “zeroed” before the dive, seen as error instance A in table 1. While this does not immediately cause any significant issues, the problem does become more complicated further along in the transcoding process.

During the dive and back on shore, MBARI Video Lab professionals annotate events seen during the dive. Each annotation is marked with the tape time code for later reference. Along with the classification of each animal, other bits of data from the ROV at the time of the annotation are also recorded (temperature, salinity, camera direction,

etc.). Once the tape is completely annotated, the associated data can be pulled from VARS. Using the VARS data, one can pinpoint exactly where in the dive specific events or animals of interest can be found without needing to analyze the entire video, saving time and unnecessary data accumulation. Users can access dive information, sort data, and select desired tapes for analysis (Schlining & Stout, 2006). In order to use the tapes for automated detection and classification purposes, they must be transcoded to a digital file, where error instance A was identified.

A dive video is placed into a Panasonic Digital HD Video Cassette Recorder and copied as a digital file using Black Magic Media Express for MAC. One aspect of the data bottleneck may be alleviated at this point by moving away from physical tapes to an all-digital file source. Currently the copying of the video is done in real time, so an hour-long tape will take an hour to copy. The MBARI video archive presently houses 29,772 tapes equaling over 22,000 hours of video that would need to be copied (Kuhn, personal communication, 2016). Once a video is copied, the result is a QuickTime (QT) video that can then be used for event detection analysis. However, the QT movie does not retain the tape time code associated with the annotations found in VARS and instead has a start time of zero with the format of HH:MM:SS. While this tends to work fine for a “zeroed” tape 1 of a dive, it does create a mathematical step to convert the VARS annotation times associated with a non-zeroed tape. Brian Schlining of MBARI created an “add-runtime” app that was used to quickly convert the tape time code in VARS to match the start of the QT movie. By first converting the VARS annotation times, smaller sections of the QT movie could be used for detection and animals of interest could be isolated. While the new runtime app was successful in rectifying error instance A, error instance B was not as quickly resolved.

During the transcoding process of digital tape to digital files with Black Magic Media, the user must click “capture” on the computer-screen while simultaneously hitting play on the VCR an arms stretch away. While some converted VARS time codes matched the QT movie’s time codes, others were off by as much as 45 seconds. In most cases, an acceptable level of error was found to be +/- 1 to 2 seconds between a VARS annotation and the QT movie. Though the animal of interest may have been annotated in VARS for a particular frame, the animal may still be visible for another few seconds

depending on the speed of the ROV. In order to get closer to a 1 to 2 second window of time, the VARS tape time codes had to be realigned to match the first frame of the QT movie. By moving through the process backwards, the QT movie was used to identify the exact starting frame of the dive video. The tape time code associated with this exact frame was then used to restore the appropriate VARS time code in the runtime app. It was found that error instance B was potentially caused by a delay between pressing both buttons or possibly from an error message that displayed while naming the digital file. A procedure manual was created to standardize the process of transcoding and resolved error B by providing consistency between the VARS annotations and the transcoded files. Most digital files were ready for detection analysis after these modifications; however, an error was found in tapes that were copied from older standard definition tapes.

Error instance C was found in tapes recorded from ROV Tiburon that were recorded on standard definition (SD) and later moved to high definition (HD). While the SD tapes were not used in this analysis, they may be used in the future and a quick solution was discovered to correct the associated time codes. Again, Brian Schlining created an alternate time code application to convert the time codes, but experience showed that there were minor fluctuations in the times and a more accurate procedure was needed. Error C arose due to the VARS annotations being entered using SD tapes, which generates an annotation time code associated with that SD tape. Later, the SD tapes were copied to an HD tape with a different time code than the VARS annotations associated with the SD tape. To make matters worse, the HD tape was then copied to a QT movie with the same error associated with error instance B. With an extra layer of human error, the times were inaccurate by over a minute for some annotations. By collecting the original SD tapes, lining up the first frames of the QT movie with that of the SD tape and using this time as the zero-time for the runtime app, a more accurate annotation time code was created. The procedure for altering the runtime app is also included in the procedure manual.

Solutions to Sources of Error in Collection		
Error Instance	Error	Solution
A	Tape time code and associated VARS annotations offset from QT movies	Add-runtime app created by Brian Schlining was used to convert the tape's time code to match the QT movie's time code in instances where the dive video did not start on 00:00:00:00
B	Delay in time code and QT movie start due to buttons not being pushed simultaneously	Procedure created to insure most accurate start times are seen between the dive video and the QT movie.
C	Alternate time code variations in SD tapes that were transferred to HD	Original SD tapes were collected, rather than using the HD tapes, and the first frame of the dive video that aligned with the first frame of QT movie was identified. The time code associated with this frame was then used with the Add-runtime app to realign the annotation times

Table 1: Sources of error in collection were first identified and then corrected. In some cases, multiple sources of error occurred.

## Video Selection

Videos were collected from both benthic and mesopelagic dives. Three benthic organisms were selected based on varying levels of abundance (high, medium, and low), scientific importance (Integrated Time Series), and location (Station M in Monterey Canyon, where all three organisms could be found in each dive) (Smith, Ruhl, Kahru, Huffard, & Sherman, 2013). These three organisms (*Peniagone*, *Scotoplanes*, and *Echinocrepis*) provided enough variation and scientific interest that the bulk of the project was spent analyzing their detection outcomes. The three mesopelagic organisms (*Nanomia*, *Eusergestes*, and *Chiroteuthis*) were also chosen based on their decreasing levels of abundance, but time did not allow for detailed detection analysis. In total, dives were analyzed from August 1999 to June 2015, chosen for the greatest abundance of organisms per dive. Due to the effects of climate change and upwelling, most of the videos collected were from 2012 when all six species were found and studied in greatest detail, as seen in figure 1 (Smith et al., 2013) Half the videos were analyzed by a partner group (results pending) and the other half was analyzed at MBARI.

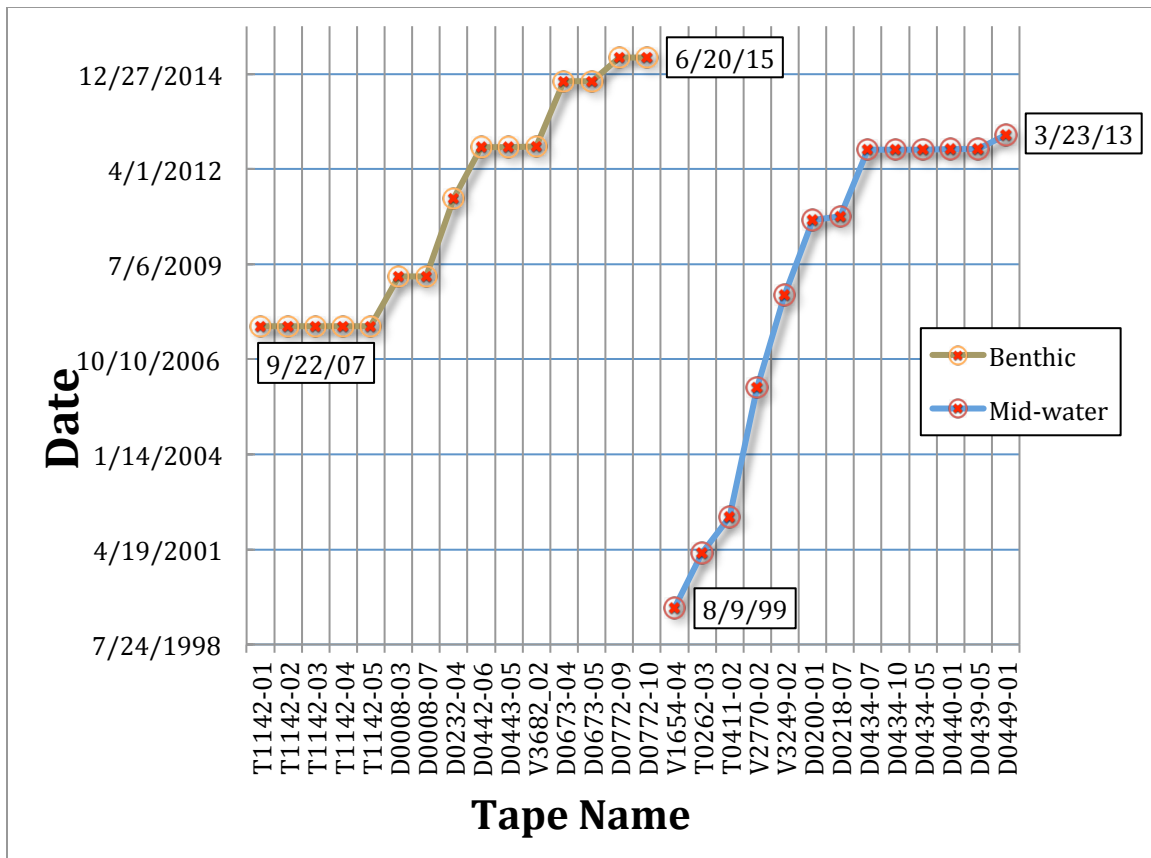


Fig 1: Tapes analyzed for both benthic and mid-water dives compared by dive date. The first letter of the tape name indicates the ROV (Tiburon, Doc Ricketts, & Ventana) and the numbers indicate dive and tape number.

## ROVs

During the times chosen for video collection, MBARI maintained 3 different ROVs. The ROV Tiburon was equipped with a WVE550 3-chip CCD (625i50, 752x582 pixels) camera with a DVW-A500 Digital BetaCam VTR for video recording in standard definition and did not produce images of high enough quality to be successful with automation (the success of future automation may allow these videos to be used). The majority of the videos were split between the Doc Ricketts, a 6'Wx12'Lx7'H Electro/Hydraulic ROV mounted with an Insite HDTV camera with 10x zoom and the Ventana, a 5'6"Wx10'Lx7'3"H Electro/Hydraulic ROV mounted with one Ikegama HD camera with HA10x5.2 Fujinon Zoom lens. Both ROVs were linked to their respective ships over a fiber optic link and video was captured using a Panasonic AJ-HD2000 high definition recorder capable of recording to digital tape and in situ annotation capabilities

with onboard VARS video capture system directed from RGB Sony Feed, HD-SDI capable. Each ROV was also equipped with scientific CTD packages to measure depth, temperature, and salinity associated with each dive and Harbor Branch Spatial lasers that can be used for species size estimations (Vehicle Technology, n.d.).

## Video Lab and Initial Detection

Once the D-5HD tapes were collected for processing, the videos were transcoded to digital files using Black Magic Media Express, converting digital tape to a QT movie ProRes 422 HQ file in 1920x1080 pixels at 29.97 frames per second. The frames are then transcoded to PPM/Netpbm color images using FFmpeg software. From here, the images are sorted using AVED software where the process of detection begins.

First, the images must go through a round of segmentation to pull the foreground from the background. All videos analyzed were from moving ROVs, so fixed camera algorithms could not be used. Variations in lighting conditions, created by the ROV as it travels across the ocean floor, create luminance gradients that affect contrast-based algorithms such as the ones used in our analysis. However, these effects are constant over the course of a dive and can therefore be removed via background subtraction (Edgington et al., 2006).

Next, the segmented frames are analyzed for saliency. Using a model similar to that of saliency-based attention in humans, each frame goes through a winner-takes-all neural network to identify objects with the highest level of saliency. Saliency, as defined in neural biology, is a property of an item (object, person, pixel, etc.) to have a state or quality that stands out relative to its neighbors. By standing out, the item can be referred to as interesting. Aspects of interest, or saliency parameters, can be selected within the AVED software for the analysis of our transcoded video images. The initial images were analyzed for this saliency based on perceived edges, contrast, and flicker (a measure of fluctuating intensity over time) (Edgington et al., 2006; Itti, Koch, & Niebur, 1998; Cline, personal communication, 2016). An object that registers with high enough saliency will then be tracked using Kalman filters to estimate its location in the next frame. If, an object can be tracked for a pre-set number of frames, it is marked as “interesting” and an image of the segmented object is then cropped out (Edgington et al., 2006; Faragher,



2012). To keep up with the computing demand of analyzing and generating these images, the videos were computed in 15 second batches on MBARI virtual machine appliance machines using a workflow management software, HTCondor (Cline, personal communication, 2016).

## Clip Selection

The process of detection highlights the importance of having the correct VARS time code associated with the QT movie being analyzed. Benthic video clips were selected by isolating events where *Echinocrepis* was found in highest numbers. Since *Echinocrepis* is found in such low abundance (41 annotations in the analyzed clips), it was important that all instances of this animal be analyzed. However, in an effort to limit the amount of data generated from each detection run, clips were narrowed down to be approximately two minutes long while capturing as many of all three benthic organisms as possible. Considering the initial sources of error created timing offsets as large as 45 seconds, had the time codes not been adjusted, these detection runs may have analyzed clips that did not contain the organisms of interest. To highlight the importance of the time codes further, had the images then gone on to classification with the wrong time code, the classification algorithm would have been trained unsuccessfully on images ranging from anything 45 seconds plus or minus the intended target. Once the videos have been transcoded and analyzed with AVED, the cropped detections were sorted for taxonomy and uploaded to Google Photos for use as training images in future neural networks.

## RESULTS & DISCUSSION

### General Overview

In all, 56 minutes and 39 seconds of video was analyzed. This generated 45,144 cropped images of everything that was detected. After these images were sorted and classified, 891 were added to the training library. A detailed analysis of this daily workflow for the benthic dives can be seen in table 2.

### Benthic

Analyzed video sections returned varying levels of success depending on marine snow density and lighting conditions. However, as seen in table 2, most sections analyzed returned an average of 9 cropped detection photos every second. Tracking was implemented to not only identify salient objects, but to also generate 3 images per detection. These images were then reviewed for the highest quality and the best image per organism was selected for use in the training library. One image per organism was used to ensure training images contained only unique examples.

Benthic Video Analysis: Initial Run				
Video Name	Time Analyzed (Minutes:Seconds)	Images Generated	Images/Second	Images Added to Library
D0008_03HD	4:30	1,915	6.9	14
D0232_04HD	7:00	3,881	9.2	24
D0442_06HD	1:31	831	9.13	13
D0443_05HD	14:17	7,666	8.9	331
D0673_04HD	1:45	1,381	7.9	55
D0772_09HD	13:45	9,310	9.3	291
Total	42:55	29,170	8.6	728

Table 2: High animal abundance sections of dive videos were selected for analysis with event detection software. The sections of analysis are broken down into the amount of images generated, images created per second, and the amount of images of high enough quality to be added to the training library.

High abundance animals proved to be the most successful with the current saliency parameters. The highly abundant *Peniagone sp.* was only detected in the video images 38.8% of the time; however, this generated 287 images that were added to the library. For the purposes of our collection, this was good enough. It is important to note that detections were not intended to find every animal in the video. The saliency parameters were set in such a way that only the “best” images of each animal would be returned. Saliency parameters can be adjusted to improve detection, but the amount of generated images and associated data will also increase. Due to the limits of manual sorting and supervision, this level of saliency was considered best case. However, the

images generated of *Peniagone* sp. appeared to show a bias towards species of darker color and higher contrast.

There are six species of *Peniagone* described in the VARS annotations and as seen in table 3, they can be divided into two main groups (“translucent” and “red”). The “red” group was detected 60.9% of the time, while the translucent group was only detected 19.5% of the time. This suggests that color may be playing an important role in detection. Since *Peniagone* was found in such high abundance, it was not necessary to re-analyze the video with color saliency filters. However, the low abundance *Echinocrepis* also displays a color bias in detection and required tuning of the algorithm.

*Echinocrepis rostrata* was annotated 41 times in VARS for the video sections that were analyzed. With a detection percentage of 80.5%, it appeared this animal was being detected more successfully than *Peniagone*. After review of the cropped images, it was determined that the AVED software was consistently missing the brown adolescent stage of *Echinocrepis*. Table 4 describes how the algorithm was tuned to improve detection. Smaller sections of video were now







<i>Peniagone</i> sp. Groups		
	<i>vitrea</i>	<b>“Translucent” Group</b> 394 Annotations in VARS 19.5% Detected 77 Images added
	<i>sp. 1</i>	
	<i>sp. 2</i>	
	<i>sp. A</i>	<b>“Red” Group</b> 345 Annotations in VARS 60.9% Detected 210 Images added
	<i>gracilis</i>	
	<i>papillata</i>	

Table 3: The six species of *Peniagone* are split into two main groups based on color and contrast. Detection percentages and the number of images added to training library are also shown.


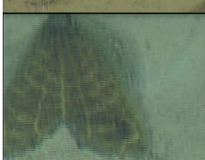

<i>Echinocrepis rostrata</i> Variations		
	White – Yellow Juvenile	41 Annotations in VARS Low Abundance 80.5% Detection rate  Algorithm Tuning: Increased Detection Size Color Saliency Channel  Data Increase: 9 pic/sec → 12
	Tan – Brown Adolescent	
	Dark Purple Adult	

Table 4: Three different life stages, number of annotations, detection rate, algorithm tuning, and data increase associated with *Echinocrepis*.

analyzed to only include the animal that was missed. In these re-analyzed video sections, all unseen instances of *Echinocrepis* were detected, but there was a slight increase in the amount of images created. There were more non-animal detections and the pictures per second rate increased from 9 pic/sec to 12 pic/sec.

The medium abundance group, *Scotoplanes globosa*, was found to have a detection percentage of 64% and resulted in 57 high quality images for the photo library. However, a high number of photos did not include large portions of the animal. This particular species has translucent skin that blends into the sand behind it, but because of the headlights on the ROV, a large dark shadow can be seen under the body. While the legs and dark shadow create high contrast with well-defined edges, the upper half of the animal fails to segment. Cropped images returned without critical distinguishing features found on the upper portions of the animal and could not be used for the training library. To combat the failure to fully segment, a Hough-based tracking filter was applied. Hough-based tracking allows for improved detection of non-rigid objects and uses back-projection to segment detected objects from their background more precisely (Godec, Roth, & Bischof, 2011). Additional algorithm tuning included increasing detection frequency from every second to every half second and saliency voltage was decreased from 4 mv to 2mv.

The detection rates were improved to include all unseen *Scotoplanes*, but this resulted in a massive increase in the number of images generated (9 pic/sec to 114 pic/sec). The results for before and after implementation of the Hough-based tracker can be seen in image 1. Image 1A shows the initial detection with focus on

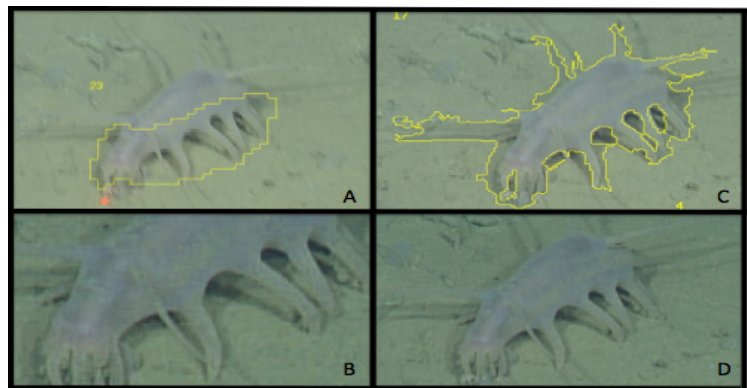


Image 1: Image A shows the initial detection limits seen for *Scotoplanes globosa* and the generated cropped image is seen in B. Image C represents the improved segmentation with the Hough-based tracker which produces a fully cropped image in D.

the high contrast legs of the animal. The detection is then cropped around the perimeter of the yellow segmentation outline and produces the image seen in 1B. While it is not yet clear if this image captured enough distinguishing elements of the animal to be successful

in future automation, the long appendages on the upper portion of the animal are critical to differentiating this particular species from the smaller *Elpidia* seen in the same environment (L. Kuhnz, personal communication, July 15, 2016). Image 1C represents the segmentation seen after the implementation of the Hough-based tracker. Not only does the tracking filter give a better true segmentation outline of the animal, it also improved the segmentation of the previously un-segmented portions of the upper half of the animal.

### Mesopelagic

Time did not allow for a detailed review of the images generated for the mesopelagic dives, but some key differences from the benthic group should be outlined. First, there are a number of distinctions in how the background is computed, the segmentation, background computation, and Kalman filter parameters. There is also no longer a need to mask the lasers seen in benthic analysis. Another key difference is the movement of animals in mesopelagic dives compared to the benthic. Since the animals we were attempting to detect were swimming through frames, tracking length was not limited, producing more images per detection. Edges were easier to detect against the dark black waters in the mesopelagic; however, there was far more potentially salient marine snow to compete with than in benthic dives.

Upon review of detection images it was seen that the animals of interest were often motion blurred or were too far away to return high quality images. Dive videos were then selected that contained video sections where MBARI dive professionals collected animals of interest. In order to sample specimens, the ROV must slow to a stop in front of the animal and the camera will generally zoom in for detailed observation of the animal. By using these videos, high quality images of *Chiroteuthis calyx* were collected and sorted. However, unlike the benthic dives, the mesopelagic dive images do not contain only unique examples. These images were chosen for different angles, lighting conditions, and distances of each individual animal.

## Productivity

Table 5 outlines the number of images sorted for both benthic and mesopelagic dives, those that were added to the training library, unique category based on taxonomic class, and the daily productivity for 30 workdays. “Sorted images” refers to the number of images cropped by analysis and only the best of the best were used for the training library. Tables 6 and 7 outline the unique categories and the number of images each contains in the new training library.

45,144 Sorted Images/891 Photos Added to Training Library		
Video Type	Benthic	Mesopelagic
Sorted Images	29,170	15,974
<b>Added to Photo Library</b>	<b>728</b>	<b>163</b>
Unique Categories	23	6
Productivity	1,505 sorted/day with 30 added/day	

Table 5: Daily productivity of 30 workdays split into benthic and mesopelagic dives. Sorted images were those that were cropped from video and the best of the best were added to the library.

Benthic Photo Library			
Category	Images	Category	Images
<b>Peniagone</b>	<b>287</b>	Umbellula	6
Benthocodon	133	Abyssocucumis	5
Elpidia	86	Crinoid	4
<b>Scotoplanes</b>	<b>57</b>	Bathypheilia australis	3
Sponge	36	Cystechinus	3
<b>Echinocrepis</b>	<b>33</b>	Hexactinellida	2
Psamminidae	18	Munidopsis	2
Epizoanthus	18	Paropsurus	2
Munnosurus	11	Striatodoma	2
Cystocrepis	10	Coryphaenoides	1
Pennatula	7	Fariometra	1

Table 6: The benthic photo library contains 23 unique categories (*Ophiuroidea* not included) and the number of images in each category is listed.

Mesopelagic Photo Library	
Category	Images
Chroteuthis	122
Eusergestes	13
D. gigas	8
Nanomia	8
Aegina	7
Jelly	3

Table 7: The mesopelagic photo library contains 6 unique categories and the number of images in each category is listed.

## **CONCLUSIONS/RECOMMENDATIONS**

The current detection procedures in place at MBARI are highly successful in analyzing video, but limitations remain on sorting and classifying images. Procedures implemented over the course of this research provide for a workflow that can be used and improved upon over time. The training image library also plays an important role in future automation towards classification. At the time of submitting this paper, the images were being used for neural network training in the hopes of developing a classification algorithm. By adding the ability to classify detected images, MBARI would no longer be limited by a sorting and classifying bottleneck. As advancements continue to be made in detection and classification, MBARI can begin analyzing more data from various platforms such as autonomous underwater vehicles, benthic rovers, time-lapse photography, and 24-hour video. Moving forward with the project I would like to conduct a more detailed analysis of the mesopelagic dive detections and I'd also like to compare detection success between video and still images, as still images may provide better quality images without motion blur. I hope that future success in this program creates the ability for MBARI to track and monitor ecosystems in situ and help provide insight on changes brought on by climate change and other environmental impacts.

## **LESSON PLAN**

Title: How Machines Learn: The Automated Detection of Deep-Sea Animals

Subject: Life Science

Grades 9 -12

To be completed in 1 – 2 class periods.

- 1.) Students will be introduced to deep-sea animals using the EARTH lesson plan, “Observing Deeply” authored by Elizabeth Rogers 2006 (<http://www.mbari.org/observing-deeply>)
- 2.) Students will discuss real world instances where they use machine learning (Facebook, self-driving cars, etc.). In groups, students will ENGAGE in a brainstorming activity to discover other ways this technology could be used to improve their lives.

3.) The conversation will then be focused on how this technology is used in marine biology (automated detection). Students will learn about 3 specific organisms (Peniagone, Scotoplanes, and Echinocrepis). To do this, students will be asked to EXPLORE Wikipedia, google, the MBARI Deep-sea guide, and any other resource to find as many images as they can in a short 10 minute time limit. Students should focus on what makes each organism unique.

3.) Students will now be given a chance to EXPLAIN these three animals as best they can as a group.

4.) Now that students have their 3 most important characteristics, they will use their phones or laptops to play the first 15 images of Kahoot! And ELABORATE on what they've learned. They will be "trained" just like a neural network by being shown images of different animals, paying special attention to the 3 characteristics they picked. Link found below.

5.) Students will now be EVALUATED by looking at more unseen images in the rest of Kahoot! (much like a neural network would) and see how many species they can positively I.D.

### **Unique Research Connections**

Coming to MBARI, I had zero experience in marine biology. One of the things I had to learn before moving forward with the machine learning is what do the animals we're trying to detect even look like? I had to become a species "expert" and it was really difficult to tell some of them apart. Some animals like *C. calyx* and *N. bijuga* look a lot alike, but are very, very different animals. The aspects of these animals that helped me tell them apart directly influenced the filters we selected for identification. The assumption being, if I can't tell it apart, neither can the computer. While this isn't necessarily always true, it is a good place to start. Machine learning is a growing trend in a lot of the technology we use today and young children should be introduced to it now. Machine learning is eerily similar to human learning and I think this provides a fun opportunity for students to see that.



### **Prior Student Knowledge**

Computer science and marine biology knowledge is not required. However, students should have a relatively good understanding of species variation, general concept of food webs, and basic computer skills.

### **Student Objectives**

Students will be able to identify the differences in mid-water and benthic ecology and what roles they play in a total marine food web. Students will have a basic understanding of machine learning and the role it plays in everyday technology (smartphones!), and how animals have adapted for mimicry

### **3 NGSS Standards Addressed**

#### **1. Asking Questions and Defining Problems**

Students will need to identify why machines may struggle to differentiate between similar looking animals. What filters would you use? Why would color be a poor choice under water? When would color be beneficial? How can you tell where something lives just by the way it looks?

#### **2. Planning and Carrying Out Investigations**

After selecting their 3 most important characteristics, students will get an idea of how fast the identification must be and test their own accuracy. Students will then be given an opportunity to try again after seeing the first set of Kahoot! pictures before hand. Was it easier or harder? What 3 characteristics were best? Did you use the same filters for both sets of pictures?

#### **3. Constructing Explanations and Designing Solutions**

Why were some filters successful and other were not? What advantage is there for an animal to be “hard to detect?” Is this the best way to detect animals? What other ways could we efficiently I.D. animals automatically? What has a better representation of animals, a moving ROV or a stationary float or sinker?

### **Suggestions for Special Adaptations (ELL, Special Needs, etc.)**

Most of the automated video detection is solely reliant on visual aspects, but there are other ways to detect species. Sound can be used for certain whales and dolphins and the audio recordings can also be used in machine learning. While some students may struggle with the machine learning concepts, the underwater footage brings an unseen world of life into the classroom that all students would love to see. ELL students can be guided through the VARS annotation system with teacher assistance to positively I.D. species as the ROV takes them across the ocean floor and Kahoot! Can be played in groups.

### **Formative Assessments**

Using Kahoot!, images of different animals will pop up on the screen and students can vote on what they think the animal is (Other, Peniagone, Scotoplanes, Echinocrepis). Kahoot! Will automatically show which students are getting the answers right or wrong and students will be given an opportunity to justify why they picked each answer. Just like a neural network, they will get better at identifying the animals.

Kahoot! Link = <https://play.kahoot.it/#/k/97d60834-106c-484b-9348-c2104f818d0b>

### **Summative Assessments**

Students will be given short answer questions to guide them towards considering concepts they may have overlooked. This lesson is more about exploration and discovery, so a lot of their responses will be opinions. General concepts may be graded, but even the experts haven't figured out the best filters or ways to approach the problem of automated detection in underwater video.

### **Possible Open-Ended Questions**

Were there any species mimics?

What does each of the three animals look like? What characteristics stand out?

What are the three most important filters (color, shape, texture, etc.) the computer should select for when trying to detect these animals?

Could you use the same filter for all three or would you have to change the filter for each animal?

What is more important; increased efficiency or accuracy? Why?

How many did you get right? Which animals did you confuse for others (species mimicry)?

Would you keep the same 3 filters?

Would it have been easier with more images? Would computers also do better with more images?

Why are some photos so blurry, based on your filters, would this be a problem?

Why would color be a poor choice under water?

When would color be beneficial?

How can you tell where something lives just by the way it looks?

### **Step-By-Step For Teachers**

1. Students will follow the EARTH lesson plan for an introduction to deep-sea creatures. Lesson plan can be found at <http://www.mbari.org/observing-deeply>
2. Facebook features a face recognition algorithm to recognize when posted photos contain people. HOW?! What makes a face unique? This lesson is not intended to teach students about computer science, but rather introduce them to the idea of what computers are capable of and how we can use this technology.
3. Have the students break into groups of 4 to think of how having a computer that could see like humans, would be useful (guide struggling students to ideas like cameras that can recognize weapons in a bank and call the police, identify important people in high security areas, medical applications to look for disease signs, etc.).
4. Groups will present their ideas of future automated detection applications
5. After groups present their ideas, guide them to using this technology to detect fish in marine biology as seen in the videos from the EARTH lesson plan. Why would this be helpful? Ecological monitoring, species discovery, etc.
6. Students will now be instructed to either “program” a futuristic computer or pretend THEY are futuristic cyborg biology computers specifically

programmed to detect fish! In their programs, they can only use 3 filters for detection (color, shape, size, shine, movement, or anything their imaginations can come up with).

7. Students will use the internet to explore images of *Peniagone* sp., *Scotoplanes* sp., and *Echinocrepis* sp. Students may also be guided through images of these organisms if more appropriate. These cyborgs will be allowed to explore as many photos as they can in 10 minutes.
8. After 10 minutes, students will play Kahoot! In groups of 4 or they can play individually if everyone has access to the internet.
9. Only play the first 15 images and pause the game. Kahoot! Will keep track of how well the students are doing. When students get answers wrong, review the image discuss why they were wrong. Was it a species mimic? Was it too blurry? Do they not know the difference between the animals? Use this interaction to gauge how they're doing. The issues that arise during this part are the same issues that arise in machine learning. Did the cyborgs not see enough images to begin with? Show them more, just like a neural network!!
10. After students/cyborgs have seen more images, play the rest of Kahoot! To see if they do any better. The hope is that as they see more images, they'll do better and begin learning the minor differences in organisms. Continue to relate their experience back to that of how computers learn.
11. Once Kahoot! Has finished, hand out the final summative assessment test. This is an open ended question/written response assignment to guide students towards deeper questions and applications of the technology they have just learned about.

## **ACKNOWLEDGEMENTS**

I would like to thank my mentor Duane Edgington for his support, patience, and reassurance that I would do fine this summer. I would also like to thank Danelle Cline for allowing me to be creative and doing all the hard work of running my video selections, suggesting ways to fine-tune the algorithms, and then helping me understand how this all

works. Thank you to George Matsumoto and Linda Kuhn for making the transition from home to MBARI and back as easy and stress-free as possible. Thank you to the summer interns and all the staff at MBARI for the encouragement and inspiration. Thank you to my wife and daughter, I couldn't do it without you. Thank you to my dad who has always been by my side and made every dream I've ever had possible. Finally, thank you to Dr. Kelly McDonald for pushing me as a student, inspiring me as a future teacher, and pushing me to try new things.

## References:

- Edgington, D. R., Cline, D. E., Davis, D., Kerkez, I., & Mariette, J. (2006). Detecting, Tracking and Classifying Animals in Underwater Video. *Oceans 2006*. doi:10.1109/oceans.2006.306878
- Faragher, R. (2012). Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation [Lecture Notes]. *IEEE Signal Process. Mag. IEEE Signal Processing Magazine*, 29(5), 128-132. doi:10.1109/msp.2012.2203621
- Godec, M., Roth, P. M., & Bischof, H. (2011). Hough-based tracking of non-rigid objects. *2011 International Conference on Computer Vision*. doi:10.1109/iccv.2011.6126228
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Trans. Pattern Anal. Machine Intell.*, 20(11), 1254-1259. doi:10.1109/34.730558
- Schlining, B., & Jacobsen Stout, N. (2006). MBARI's Video Annotation and Reference System. Presented at the Proceedings of the Marine Technology Society / Institute of Electrical and Electronics Engineers Oceans Conference.
- Smith, K. L., Ruhl, H. A., Kahru, M., Huffard, C. L., & Sherman, A. D. (2013). Deep ocean communities impacted by changing climate over 24 y in the abyssal northeast Pacific Ocean. *Proceedings of the National Academy of Sciences*, 110(49), 19838-19841. doi:10.1073/pnas.1315447110
- Vehicle Technology. (n.d.). Retrieved August 02, 2016, from <http://www.mbari.org/technology/emerging-current-tools/vehicle-technology/>
- Walther, D., Edgington, D., & Koch, C. (2004). Detection and tracking of objects in underwater video. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. doi:10.1109/cvpr.2004.1315079