



**CALIFORNIA
ENERGY COMMISSION**



Energy Research and Development Division

FINAL PROJECT REPORT

Lowering Costs of Benthic Ecological Surveys Through Artificial Intelligence to Inform Offshore Renewable Energy

Gavin Newsom, Governor
September 2020 | CEC-500-2020-064

PREPARED BY:

Primary Authors:

Lubomir Stanchev
Benjamin Ruttenberg
Alexandra Wolman

Cal Poly Corporation
California Polytechnic University San Luis Obispo
1 Grand Ave, Building 15
(CPCorp) Phone; <https://www.calpolycorporation.org/>

Contract Number: EPC-17-029

PREPARED FOR:

California Energy Commission

David Stoms, Ph.D.

Project Manager

Jonah Steinbeck, Ph.D.

Office Manager

ENERGY GENERATION RESEARCH OFFICE

Laurie ten Hope

Deputy Director

ENERGY RESEARCH AND DEVELOPMENT DIVISION

Drew Bohan

Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees, or the State of California. The Energy Commission, the State of California, its employees, contractors, and subcontractors make no warranty, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission, nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

ACKNOWLEDGEMENTS

The researchers thank the California Energy Commission for all their support in developing this technology. The researchers also acknowledge the general support of Amazon Web Services (AWS), which provided \$80,000 in AWS credits. This work would also not be possible without the help of Monterey Bay Aquarium Research Institute (MBARI) which provided 50 hours of breathtaking, high-definition underwater video transects in addition to crucial guidance from MBARI data engineer and video annotators: Danelle Cline, Jacqueline (Jacki) Long, Duane Edgington, and Lonny Lundsten. The PIs would also like to personally thank our Technical Advisory Committee: Chris Potter, Donna Schroeder, Shari Matzner, Franz Kurfess, Nick Osturbur, Sam Blakeslee and Danelle Cline for their time and insight. Finally, the researchers thank Cal Poly San Luis Obispo graduate student Hanson Egbert and undergraduate students Viet Nguyen, Ishaan Jain, Samantha Gunzl, Jacob Territo, Kyle Maxwell, Kyaw Soe, Trace Rainbolt, Kevin Label, Stephen Hung, and Justin Cho, from the Computer Science and Software Engineering Department who developed the website software, Elly Halladay from the Graphic Design Department for designing the website and Emma Barton, Kellen Breitenbach, Allison Matthews, and Sierra Smith, from the Biological Sciences Department, who annotated the videos.

PREFACE

The California Energy Commission's (CEC) Energy Research and Development Division supports energy research and development programs to spur innovation in energy efficiency, renewable energy and advanced clean generation, energy-related environmental protection, energy transmission and distribution and transportation.

In 2012, the Electric Program Investment Charge (EPIC) was established by the California Public Utilities Commission to fund public investments in research to create and advance new energy solutions, foster regional innovation and bring ideas from the lab to the marketplace. The CEC and the state's three largest investor-owned utilities—Pacific Gas and Electric Company, San Diego Gas & Electric Company and Southern California Edison Company—were selected to administer the EPIC funds and advance novel technologies, tools, and strategies that provide benefits to their electric ratepayers.

The CEC is committed to ensuring public participation in its research and development programs that promote greater reliability, lower costs, and increase safety for the California electric ratepayer and include:

- Providing societal benefits.
- Reducing greenhouse gas emission in the electricity sector at the lowest possible cost.
- Supporting California's loading order to meet energy needs first with energy efficiency and demand response, next with renewable energy (distributed generation and utility scale), and finally with clean, conventional electricity supply.
- Supporting low-emission vehicles and transportation.
- Providing economic development.
- Using ratepayer funds efficiently.

Lowering Costs of Benthic Ecological Surveys Through Artificial Intelligence to Inform Offshore Renewable Energy is the final report for the Lowering Costs of Underwater Biological Surveys to Inform Offshore Renewable Energy project (Contract Number EPC-17-029) conducted by Cal Poly Corporation. The information from this project contributes to the Energy Research and Development Division's EPIC Program.

For more information about the Energy Research and Development Division, please visit the [CEC's research website](http://www.energy.ca.gov/research/) (www.energy.ca.gov/research/) or contact the CEC at 916-327-1551.

ABSTRACT

As California works to increase its renewable energy portfolio and explores opportunities to develop offshore renewable energy, the need grows for pre-construction biological surveys and post-construction monitoring of deep-water benthic environments. Underwater video is a powerful tool to facilitate such surveys; however, interpretation of video surveys by trained annotators into usable data is costly and time consuming. Emerging technologies have improved automated analysis of underwater video, but many still lack accuracy or accessibility. To address these challenges, the researchers developed DeepSeaAnnotations.com, a free and open-source, web-based software that enables users to (1) watch survey video and save annotations of organisms as image, species, and spatial-temporal data; (2) track movement of annotated organisms forward and backward in time to automatically generate video clips of organism movement and additional image data; (3) verify quality of annotations and tracking videos; (4) create and train artificially intelligent neural network models with custom batches of annotation data; and (5) use the models to automatically annotate unwatched videos. Further, the project demonstrated that it is possible to transfer the workflow for annotating underwater video to a web-based cloud environment, providing significant savings in time and expense by enabling scientists to use the software from any geographic location and work in real time.

Undergraduate marine biology students used the software to interpret 50 hours of high-resolution, benthic survey video provided by the Monterey Bay Aquarium Research Institute, resulting in more than 40,000 annotations of more than 100 classifications of deep-sea, benthic species. More than 4,500 annotations were then verified for accuracy of species identification by marine biology experts. F1 scores, which measure the fraction of annotations that were correct and the fraction of objects that were correctly annotated in new videos, for five environmentally important species varied between 0.61 and 0.76. It is expected that higher accuracy could be achieved with more verified annotations for each species.

The platform from this proof of concept could be readily expanded to include more species.

Keywords: deep-sea annotations, artificial intelligence, neural networks, offshore wind energy

Please use the following citation for this report:

Stanchev, Lubomir, Benjamin Ruttenberg, and Alexandra Wolman. 2020. *Lowering Costs of Benthic Ecological Surveys Through Artificial Intelligence to Inform Offshore Renewable Energy*. California Energy Commission. Publication Number: CEC-500-2020-064.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	i
PREFACE	ii
ABSTRACT	iii
EXECUTIVE SUMMARY	1
Background	1
Project Purpose.....	1
Project Approach.....	2
Project Results	2
Technology/Knowledge Transfer	3
Benefits to California	3
CHAPTER 1: Introduction	5
Renewable Energy and Environmental Analysis	5
Artificial Intelligence: Facilitating Data Collection	6
Objectives and Organization of the Report.....	7
CHAPTER 2: Project Approach	9
Five-Phase Software Development Plan	9
Phase 1: Develop Web-Based Video Annotation Interface	9
Phase 2: Training Neural Network Models	10
Phase 3: Annotation Data Quality Control Tool	11
Phase 4: Automatic Bounding-Box Annotations Displayed in Videos of Organisms.....	11
Phase 5: Software Displays Video Metadata and Ecological Metrics	12
CHAPTER 3: Project Results.....	13
Summary.....	13
Project Results	13
Phase 1: Develop Web-Based Video Annotation Interface	13
Phase 2: Train Neural Network Models.....	14
Phase 3: Annotation Data Quality Control Tool	15
Phase 4: Automatic Bounding-Box Annotations Displayed in Videos of Organisms.....	16
Phase 5: Software Displays Video Metadata and Ecological Metrics	17
CHAPTER 4: Technology/Knowledge/Market Transfer Activities.....	18
Current Technology Availability	18
Dissemination of Project Deliverables	18

Proof of Concept for Future Work.....	18
CHAPTER 5: Conclusions/Recommendations.....	19
Lessons Learned	19
Conclusions and Future Work.....	21
CHAPTER 6: Benefits to Ratepayers	23
Technological Benefits	23
List of Acronyms	25
REFERENCES	26
APPENDIX A: User-in-the-Loop Guide.....	A-1
APPENDIX B: CNN Model Results	B-1

LIST OF FIGURES

	Page
Figure 1: Screenshot of Annotation	14
Figure 2: First Screenshot of Moving Crab	15
Figure 3: Second Screenshot of Moving Crab	16
Figure 4: Screenshot of Summary Table	17
Figure A-1: Concept Selection Interface.....	A-2
Figure A-2: Annotation COLLECTION INFO window.....	A-3
Figure A-3: Annotation Collection Interface: Selecting Users.....	A-4
Figure A-4: Annotation Collection Interface: Selecting Videos	A-5
Figure A-5: Annotation Collection Interface: Selecting Concepts.....	A-6
Figure A-6: Concept Collection Interface.....	A-7
Figure A-7: Video Collection Interface	A-8
Figure A-8: Video Annotation Statistics Interface.....	A-9
Figure A-9: Video Information Interface	A-10
Figure A-10: Video Annotation Interface: Selecting a Videos.....	A-11
Figure A-11: Video Annotation Interface: Playing a Videos.....	A-12
Figure A-12: Video Annotation Interface: Creating a Bounding Box	A-13
Figure A-13: Video Annotation Interface: Selecting the Concept for an Annotation.....	A-14
Figure A-14: Annotation Verification Interface: Selecting the Annotation Collection to Verify	A-15

Figure A-15: Annotation Verification Interface: Verifying an Annotation	A-15
Figure A-16: Annotation Verification Interface: Verifying a Tracking Video	A-16
Figure A-17: Annotation Verification Interface: Transitioning to Next Frame	A-17
Figure A-18: Reporting Interface: Selecting Sort Order	A-18
Figure A-19: Reporting Interface: Shows Number of Annotation	A-18
Figure A-20: Reporting Interface: Tree Hierarchy	A-19
Figure A-21: Reporting Interface: Example Annotation	A-20
Figure A-22: Model Interface: Creating a New Model	A-21
Figure A-23: Model Interface: Models Dashboard	A-22
Figure A-24: Model Interface: Training a New Model Version	A-23
Figure A-25: Model Interface: Training Info	A-23
Figure A-26: Model Interface: Displaying the Results of Training a Model Version	A-24
Figure A-27: Account Interface: Changing a User's Password	A-25
Figure A-28: Account Interface: Creating a New User	A-25
Figure A-29: Account Interface: Displaying User Information	A-26

LIST OF TABLES

	Page
Table B-1: Top Concepts Trained with Human Data	B-2
Table B-2: Top Concepts Trained with Human and Generated Data	B-2
Table B-3: 5,000 Annotations Per Concept, Good Data	B-3

EXECUTIVE SUMMARY

Background

California's pursuit of its goal to have 100 percent of total retail sales of electricity come from eligible renewable energy resources and zero-carbon resources by 2045 is spurring exploration of opportunities to develop offshore renewable energy capacity. In addition to the technical challenges, many unanswered questions about the environmental effects of such facilities remain. Before any projects can be approved, scientists will need to conduct pre-construction biological surveys and post-construction monitoring in the challenging marine environment. Such surveys will have to investigate what species occur in a project area and how abundant they are. Underwater video is a powerful tool to facilitate such surveys; however, identification of all the organisms in the imagery is costly and time-consuming.

Emerging technologies have greatly improved automated analysis of underwater video using a variety of approaches in artificial intelligence, but these technologies are not yet sufficiently accurate or accessible for widespread adoption in the scientific community or industries that might benefit from these tools. While there have been recent advances in automated image classification techniques, many objects and species require custom and fine-tuned processing to provide reliable and high accuracy classification capabilities for environmental baseline and monitoring tasks. Finally, these capabilities have yet to be integrated into automated annotation software that accommodates the users (known as "human-in-the-loop"), limiting the potential for significant gains in efficiency and meaningful data. As such, it is necessary to develop web-based software tools that will help humans efficiently identify objects in the underwater videos and label (annotate) them as the appropriate species, develop automated models that are "trained" to recognize these species, evaluate and improve these models, and ultimately disseminate the technology to facilitate environmental analysis assessments and ongoing monitoring.

Approaches in artificial intelligence (such as neural network modeling) have the potential to revolutionize how to conduct environmental assessments and monitoring of not only offshore wind energy projects but potentially any project that uses large amounts of video or photographic data. However, many existing tools are custom-designed and built for a single specific purpose or group of users or both. In addition, the neural network models require an enormous amount of training data to classify targets correctly, and generating this training data can be extremely time-consuming and expensive. However, advances in cloud computing have made it possible to move some tools to the cloud, giving users located in diverse geographical locations remote access and enabling researchers to collect data in real or near-real time, thus reducing the prohibitive time and expense involved previously.

Project Purpose

This project developed an efficient system to generate and use artificial intelligence models to facilitate assessments of deep-sea organisms, which will help stakeholders in California determine what sites might be suitable for offshore wind facilities. The researchers developed a simple, web-based interface that enables users to generate training data (the labels identifying species in the videos), quality check the data, develop artificial intelligence models, test these models, and provide feedback to improve the computer-generated models. To demonstrate this proof of concept, the researchers annotated a large sample of underwater

videos for a handful of common species that can be found off the California coast and ran the automated models to annotate the verification videos.

Project Approach

A team of marine biologists and computer scientists at California Polytechnic State University San Luis Obispo (Cal Poly) developed a simple, web-based interface that enables users, such as scientists and managers or energy developers, to automate the expensive and time-consuming process of reviewing information about organisms in the deep sea. This information is essential to evaluate any proposed or operating offshore wind-energy facility. The strategy is to attack several challenges at once: first, it enables users to efficiently annotate training videos, thereby reducing the costs to generate the necessary training data. Second, it enables users, including outside experts, to efficiently verify the potential training data. This is an essential step because misclassified input data will negatively affect model accuracy. Third, it helps users develop models for specific targets (such as species or “concepts” in computer science terminology). Fourth, it enables models to run on specific pre-annotated datasets, such as annotated videos that were not included in the training data. This provides instant feedback on model accuracy. Finally, it enables users to correct the errors of the model, where information is fed back to the model to improve performance. Ultimately, all of these steps could be combined to produce models with high accuracy that can be used on novel input datasets, such as videos that no user has annotated.

The Monterey Bay Aquarium Research Institute provided the videos that were annotated. Students from the Computer Science and Software Engineering Department at California Polytechnic State University wrote all the source code, trained different neural network models, used them to automatically annotate verification videos, and measured the quality of the automatic annotation. Once the annotation software was ready, a team of marine biology students used it to annotate the videos. The head marine biologist on the team trained the marine biology students to recognize the different species, such as star fish, urchins, and rockfish, and follow consistent annotation practices. These students created more than 40,000 annotations and verified several thousand annotations. Throughout the process, they continually made suggestions to the programming team for features that would improve the annotation process.

Project Results

This project has provided a proof of concept for a web-based system that enables users to annotate and verify video data, build neural network models to automatically classify species, and evaluate the F1 accuracy of the models. The F1 score is a combination of two metrics—the fraction of annotations that were correct out of all the annotations made and the fraction of objects that were correctly annotated out of all the objects that occur in the video. Achieving high F1 scores means that the algorithm is generally good at annotating objects. For the five chosen concepts, the system exhibited F1 score accuracy of between 0.61 and 0.77, which demonstrates that as a proof-of-concept, the system can be applied to real-world applications. The research team developed a website that enables a user to:

- Quickly play and annotate underwater videos.
- Create a short tracking video for each annotation that shows how an annotated concept moves in time.

- Verify the accuracy of existing annotations and tracking videos.
- Create a neural network model from existing annotations.
- Automatically annotate unwatched videos using a model that was previously created.

Using the website, more than 40,000 annotations of more than 100 different concepts were made. Of these annotations, more than 4,500 were verified for accuracy. Greater accuracy with the neural network models would be expected with more annotations and verifications for each species.

Technology/Knowledge Transfer

As a proof of concept, the technology is not ready for market, but the work has provided a clear roadmap for what is necessary to make this project operational — more investment in the user interface, more training data to train the models, more efficient neural network algorithms, and hierarchical models that can deal with targets at the species level and higher group levels. The website source code is available for anyone who would like to build on the work.

The project's GitHub page has been cloned by 58 unique visitors, including by Monterey Bay Aquarium Research Institute and the Digital Transformation Hub at the California Polytechnic State University. The webpage at www.deepseaannotations.com describes the project. The project's GitHub page is github.com/video-annotation-project. The page includes all the source code and instructions on how to install it. All the software is available to the general public under the Apache License. The videos on the website www.deepseaannotations.com are password protected because they are property of the Monterey Bay Aquarium Research Institute. However, anyone can download the source code from the GitHub repository: github.com/video-annotation-project. Once the source code is downloaded, someone with a computer science background can follow the installation instructions and launch their own website that can be used for human and automatic annotation of videos.

The project was presented twice to the technical advisory committee—October 2018 and November 2019. The committee consisted of experts in marine biology, computing, and remote sensing technology from the Ocean Protection Council, Bureau of Ocean Energy Management, Pacific Northwest National Laboratory, Amazon Web Services, Monterey Bay Aquarium Research Institute, and Cal Poly members from the Department of Computer Science and Software Engineering and the Institute for Advanced Technology and Public Policy. The project and all deliverables have also been shared with the Monterey Bay Aquarium Research Institute and the Digital Transformation Hub at Cal Poly. The results of the project were published at the IEEE International Conference on Semantic Computing.

Benefits to California

This project has provided a proof of concept of a web-based system that enables users to annotate and verify video data, build neural network models, and evaluate their performance. With further development, this technology could be used to evaluate an enormous amount of video data from potential areas of offshore wind energy development, potentially saving years of work over having humans review this information. Such savings would lower the costs and time necessary for environmental assessment, giving California stakeholders access to better information about potential environmental effects of offshore wind more quickly and cheaply. Furthermore, this technology and the framework developed have applications far beyond

deep-sea benthic marine organisms. They could be used for ongoing monitoring of a suite of other biological indicators (such as birds or marine mammals) for marine and terrestrial applications, as well as almost any target that an agency might need to monitor or evaluate from video or photo data. Future applications of this tool could extend to a range of scientific and monitoring needs for nearly any agency or industry.

CHAPTER 1:

Introduction

Renewable Energy and Environmental Analysis

Increased demand for electrical energy and concerns about the effects of climate change have prompted many governments at all levels to set aggressive goals to reduce greenhouse gas emissions and increase the supply of renewable energy. California is no exception, with the 100 Percent Clean Energy Act of 2018 (De León, Chapter 312, Statutes of 2018) requiring an increase of renewables to 60 percent by 2030 and 100 percent of electric retail sales to come from renewable energy resources and zero-carbon resources by 2045. Part of this supply will be met by solar and onshore wind (Graabak and Korpås 2016), but since both of these sources are intermittent, there are efforts to diversify energy portfolios. One response to these changes is the recent, dramatic increase in the development of commercial-scale offshore wind (OSW) energy facilities (IRENA 2016). In 2017 alone, the total installed offshore wind capacity globally rose 4,334 Megawatts (MW) to a total of 18,814 MW and is forecasted to reach 120 Gigawatts (GW) worldwide by 2030 (GWEC 2018).

Over the last few decades, the offshore wind energy industry has expanded its scope from turbines mounted on fixed platforms driven into the seafloor and standing in less than 60 meters of water (for example, Vindeby, Denmark; 4C Offshore 2017), to floating turbines moored in 120 meters of water (such as Hywind Scotland, Scotland; 4C Offshore 2018), to prospecting the development of floating turbines moored in approximately 1000 meters of water (for example, Castle Winds proposal, California; Trident Winds 2016). Major incentives to install deep-water, floating OSW facilities include reduced effects on human activities and marine ecosystems, the ability to leverage existing infrastructure and technological advancements from the offshore oil and gas industry, and access to larger and more consistent wind speeds offshore (Musial and Ram 2010, James and Costa Ros 2015, and Wang et al. 2019). However, the associated technology is still in its infancy, with few prototype turbines and mooring systems currently in use. Thus, the potential effects of these technologies on the marine environment are largely unknown.

To evaluate and issue permits for any potential OSW project, significant environmental and socio-economic analysis will be required. In addition, significant environmental monitoring will be required before, during, and after installation. For many of these factors, few data exist, to the extent that substantial effort will be needed to generate the relevant environmental data of deep-sea benthic (bottom-dwelling) organisms. Deep-sea organisms have the potential to be significantly affected by installation of OSW facilities; the facilities may require many miles of cables and enormous anchors to fix the positions of turbines. Collecting data on the abundance of key deep-sea benthic fishery and ecological species will be challenging because the great depths make accessing these areas difficult and expensive. Usually, researchers collect this information by deploying remotely operated vehicles (ROVs) that collect video transects at the relevant depths in the relevant habitats. However, identifying these organisms in the videos is difficult, time-consuming, and expensive for several reasons. There are relatively few biologists who have the appropriate expertise to identify all (or most) of the organisms present on a given video transect. The software to identify and count relevant organisms automatically is often custom-built and therefore only usable in certain locations by

certain trained observers. In addition, when there are many individuals present on a video transect, it can take an hour or more for a trained observer to process just a few minutes of video. As a result, it can take up to two years and extensive financial resources to manually process all video from a given research cruise, all of which would simply delay the evaluation of a given proposed OSW project and increase the overall costs of installation should a project be approved. As a result, there is a strong incentive to find alternative ways to process this information faster and with less cost.

Artificial Intelligence: Facilitating Data Collection

Emerging research in artificial intelligence (AI) and deep learning has demonstrated the potential to revolutionize how to conduct environmental assessments and monitoring of not only offshore wind energy projects but potentially any project that uses video or photographic data. However, many of the tools are custom designed and built for a single specific purpose or group of users or both, which greatly limits their use. In addition, the neural network models require an enormous amount of training data to classify targets correctly. Deep-sea organisms are excellent candidates for the development of web-based artificial intelligence tools. Data collection is challenging for humans, yet highly repetitive. Trained humans cue on specific visual features, such as texture, edges, and color contrast. These are the sorts of features that neural network systems can also be trained to recognize.

Generating the training data can be extremely time-consuming and expensive. However, advances in cloud computing and cloud storage have made it possible to move some tools to the cloud. Experts are often located at institutions in various locations. Collaboration would be greatly improved by using web browsers that access a common repository stored on the cloud, which enables users to access the tools and store the data from anywhere and increase the use of the tools.

Finally, development of a high-performing system would enable researchers to collect data in real time or near-real time, greatly reducing the cost of generating information on the abundance of deep-sea organisms that is critical to any environmental analysis, whether that analysis is to evaluate the location of a proposed OSW generation facility, examine the effects during construction, or monitor the response of these sensitive ecological communities during operation.

An appropriate tool should be web-based, so that users and experts can access information from anywhere. It should include an efficient way for users to generate the key training data required to develop the convolutional neural network (CNN) models because that is the most time-consuming and therefore the costliest step in using the tool for a new application. This step is called annotation, where users identify organisms they see in the videos. Users or other experts need to be able to verify (that is, quality check) the annotations that others have generated because accurate training data is essential for the CNN models to perform well. Users need to be able to then train models, ideally by choosing the species ("concepts," in computer science terms) of interest and the training data (for example, selected high-quality videos). Next, the models need to run on videos that have been annotated, but that were not included in the training dataset to build the model. These verification videos are used to assess the accuracy of the computer-generated annotations. Users need to be able to verify the annotations from the CNN models, providing additional feedback to the models. This process is repeated until models reach a satisfactory level of accuracy. Finally, models can be

applied to new datasets (that is, videos) to generate information on the abundance of the organisms that they were trained to recognize.

A successful tool can be used for a variety of different applications for evaluation or monitoring of OSW facilities. In addition to applications for deep-sea benthic organisms that this project seeks to develop, such a framework could be used to evaluate the effects of OSW development and operation on marine birds or mammals, using the exact same framework and web-based tool described here. This framework could also be used to evaluate effects of terrestrial wind, solar, or other energy facilities (Furness et al., 2013, Bailey et al. 2014, Henkel et al. 2014) on terrestrial birds or other animals.

Objectives and Organization of the Report

The main objectives of the project are:

1. Collect deep-sea underwater videos from Monterey Bay Aquarium Research Institute (MBARI).
2. Examine the videos and identify concepts that are abundant in the videos and at the same time are important from an environmental standpoint.
3. Create web-based annotation software, where the videos and data are stored on the cloud. This will enable multiple users to annotate the videos from any remote computer.
4. Use the web-based software to annotate the MBARI videos.
5. Create web-based software that allows the verification of existing annotations.
6. Use the software to verify existing annotations.
7. Create cloud-based software that creates additional annotations by tracking an object in a human-made annotation back and forward in time (that is, from the moment the object appears in the video to the moment it disappears from the video).
8. Create cloud-based software that creates tracking videos from each annotation. The videos should be around six seconds long and show the movement of an object around the time it was annotated.
9. Create cloud-based software that can display all the annotations for a particular concept, including tracking videos. This can be used to correct existing annotations and to educate the annotators about different concepts.
10. Create cloud-based software that enables the creation of collections of concepts, videos, and annotations. This will enable the user to more easily create and train models.
11. Create cloud-based software that allows the user to create a neural network model. The user just needs to specify the concepts of interest and the verification videos on which to compute the accuracy of the model.
12. Create cloud-based software that allows the user to create different versions of a neural network model. For each version, the user should specify the set of annotations that should be used for training.
13. Implement cloud-based software that trains a version of a neural network model using the associated annotation collection. The user should have the option to specify how many annotations should be used to train the model. The software should have the capability to automatically annotate the verification videos and report the accuracy for

the concepts in the model. The software should be smart enough to use only high-quality annotations. In other words, prioritize verified annotations over unverified ones, prioritize computer-generated tracking annotations from slow videos, which are generally of better quality, and prioritize computer-generated tracking annotations that are closer to the annotation, which are also generally of better quality.

14. Create cloud-based software that can be used to automatically annotate a previously unwatched video using a neural network model version. The software should provide summary statistics, including the number of detected objects from each concept and the total density in the video for each concept.
15. Create cloud-based software that facilitates reinforcement learning. In other words, a human can use the verification software to correct computer-generated annotations and then use the corrected annotations to create a new neural network model version. In this way, the computer can “learn” from its mistakes.

The researchers seek to continue development of the proof-of-concept tool described here, so that it will be useable in the future if and when environmental data collection begins to analyze the suitability of OSW deployment on California’s Outer Continental Shelf.

Chapter 2 presents the project approach. Chapter 3 shows the project results, while Chapter 4 focuses on the technology, knowledge, and market transfer activities. Chapter 5 contains the conclusion and recommendations, while Chapter 6 outlines the benefits of the project to the ratepayers.

CHAPTER 2:

Project Approach

Five-Phase Software Development Plan

To achieve both the web application and CNN model software objectives, the team divided the software tasks into five phases roughly based on our project objectives, outlined below. In practice, the phases were not entirely done sequentially, as some features faced obstacles and others became more urgent than originally planned.

Phase 1: Develop Web-Based Video Annotation Interface

The main objectives of this initial project phase were to first develop a functional video annotation interface and second to use the software to generate data and feedback on the website functionality. The initial software requirement was to develop a video player with controls that allowed the annotation of concepts.

Once the software was functional, a team of marine biologists used it to annotate the MBARI videos. These videos were heterogeneous: different depths, different resolution, different lightning, and so on. The head marine biologist on the team trained the marine biology students to recognize the different concepts and follow consistent annotation practices. Throughout the process, the team of annotators continually made suggestions for features that would streamline and improve the annotation process. Some of these features were integrated in the software.

The video control software features are:

- The ability to select a video from the left drawer. A video can be selected from the `My in Progress Videos`, `Unwatched Videos`, `Annotated Videos`, `All in Progress Videos`, or a specific video collection. Note that videos that are annotated by more than one annotator are displayed in red.
- The ability to play or stop the video, go to full screen, or go forward or backward by one second. These can be done through buttons that are displayed below the video or through keyboard shortcuts.
- The ability to change the play rate of the video from 0 (that is, video is not playing) to 4 (the video is playing four times faster than the recorded speed).
- The ability to mark the video as annotated.
- The ability to toggle the controls in the video. The controls allow users to set the volume, go full screen, download the video, or move to a specific time in the video.

The annotation control features include:

- The ability to annotate an object in the video using a bounding box that is drawn with the mouse.
- The ability to select a concept for the annotation from the concept drawer. If the concept does not exist, the user can search for it and add it to the basket.
- The ability also to drag and drop the concepts in the concept drawer to put more frequently used concepts first.

- The ability to create an annotation and mark it as “unsure” so that it can be checked later by a different annotator.
- The ability to add a comment to each annotation.

The concepts control features are:

- The ability to load the concepts from a JSON file.
- The ability to display the concepts in a hierarchy.
- The ability to search for a concept by name or ID.
- The ability to add a concept to the concept basket of the user.
- The ability to delete a concept from the concept basket of the user.

The collections control features are:

- The ability to create a collection of concepts.
- The ability to create a collection of videos.
- The ability to create a collection of annotations.

The account management features enable users to:

- View the user profile and change the current profile.
- Add a regular or administrative user.
- Find the number of annotations and verifications for a given user in a certain time period. This can be used to monitor the work of the annotators.

Phase 2: Training Neural Network Models

The software development team divided into two sub-teams to continue web application development and work on preliminary CNN model development. The objective of this project phase was to train CNN models to correctly recognize and identify five benthic deep-sea species or concepts with high accuracy.

Slow progress in creating human annotations and low quality of some of the video used to generate annotation training data created obstacles to improve CNN model performance. CNN model performance is dependent on a large quantity of high-quality annotation data. The senior marine biologist on the project verified existing human annotations in the hope of improving the model’s accuracy. Our CNN model is based on RetinaNet 50 (Lin 2017).

The main features of the CNN software are listed next.

- The ability to create a new model. When doing so, the user specifies the model name, concepts of interest, and verification videos.
- The ability to create a new model version for a specific model. When doing so, the user needs to specify the annotations to be used to train the CNN, the number of epochs for the training, and the number of annotations to use for each species. The user also has the option to select whether to include computer-generated annotations from tracking and whether to use only verified annotations.
- The ability to show the accuracy of the model version on the verification videos once a new model version is created. The verification videos are also automatically annotated.

This can help the user perform quality control and see where the model needs improvement.

- The ability to use a model version to annotate a previously unwatched video, where the result is saved to the database as computer annotations.

Phase 3: Annotation Data Quality Control Tool

The researchers originally planned to build a quality-control tool to find errors in human or AI data after completing the training of the CNN models. However, to properly assess model performance, it is important to eliminate low-quality training data. The software development team built a data quality-control interface into the web software, where the features are listed next.

- The ability to select the annotation collection that needs to be verified. The user can also choose whether to include annotations from tracking.
- The ability to switch between each annotation and the corresponding tracking video. The tracking video can be used to give the annotator more information to decide whether the annotation is correct. Both the annotation and the tracking video can be marked as correct or wrong.
- The option to change the bounding box or concept for the current annotation and seeing more information about the annotation, including video, annotator, and sea-depth of the annotation.
- The option to delete any of the annotations in the current frame or change them. Five colors are used to represent the different bounding box annotations: red for hovered over, green for verified in current collection, blue for a human annotation that is outside the current set of concepts, yellow for a computer-generated annotation that is outside the current set of concepts, and orange for an annotation in the current collection of annotations that needs to be verified.
- Memory that identifies the annotators' current place in the process and enables them to continue verifying annotation the next time they log in.

Phase 4: Automatic Bounding-Box Annotations Displayed in Videos of Organisms

The tracking software traces each human or CNN-made annotation forwards and backwards in time to create a tracking video that shows an organism object within a video. Tracking software was developed in the Phase 1 of the project to generate videos of organisms within the data report enabling users conducting quality control to view annotated organisms in motion when the video frame does not have enough clarity to identify a concept. This feature increased the efficiency and accuracy of preliminary data quality control using the report tab (before verification software was developed). In addition, tracking videos annotate organisms across multiple frames, enabling the annotators to both avoid duplicate annotations and catch organisms that were previously missed.

Phase 4 is implemented through the reporting tab, where its main features include the following.

- It first allows the user to choose the sorting order, where annotations can be sorted by any of the following criteria: Video, Concept, or Annotator. It also allows the user to select Unsure Only, or select annotations that are verified, not verified, or both.
- The user can use the tool for different reasons.
 - To check the work of an annotator by examining all their annotations.
 - As an educational tool to see all the annotations and associated tracking videos for a concept.
 - To verify unsure annotations.
- The tool has the functionality of showing the tracking video for each annotation frame and enabling the user to move or resize the bounding box, change the concept of an annotation, or delete an annotation altogether.

Phase 5: Software Displays Video Metadata and Ecological Metrics

Phase 5 displays video metadata including start and stop Global Positioning System (GPS) coordinates and ecologically relevant metrics. This includes the density (that is, the number of object occurrences for each concept per mile) and the total count of each concept of interest in a video (that is, the total number of times an object of that type appeared in a video).

CHAPTER 3:

Project Results

Summary

The website was successfully built. Some of the tasks that users can perform on the website include the following:

- Humans can annotate videos.
- Existing annotations can be verified.
- Videos, annotations, and concepts can be grouped into collections.
- Additional meta information about videos, annotations, and concepts can be stored.
- Summary statistics for each video, including abundance summary statistics for each concept, can be calculated.
- The user can learn about different concepts by examining annotations and watching tracking videos.
- The user can train models using existing annotations and use these models to annotate previously unwatched videos.
- The user can validate the quality of a model by examining the accuracy of the model on the verification videos (that is, videos that are annotated by a human, but not used to train the model).
- The user can watch computer-annotated videos that contain bounding-box concept annotations of previously unwatched videos.
- The user can examine summary statistics about the work that is done by each annotator. This includes the number of annotations and verification done by each annotator for each video and species, including aggregate counts.

The main goal of the website is to provide a tool that annotators can use to create high-quality annotations with minimal effort.

Project Results

The following sections describe the results of each of the five phases outlined in Chapter 2.

Phase 1: Develop Web-Based Video Annotation Interface

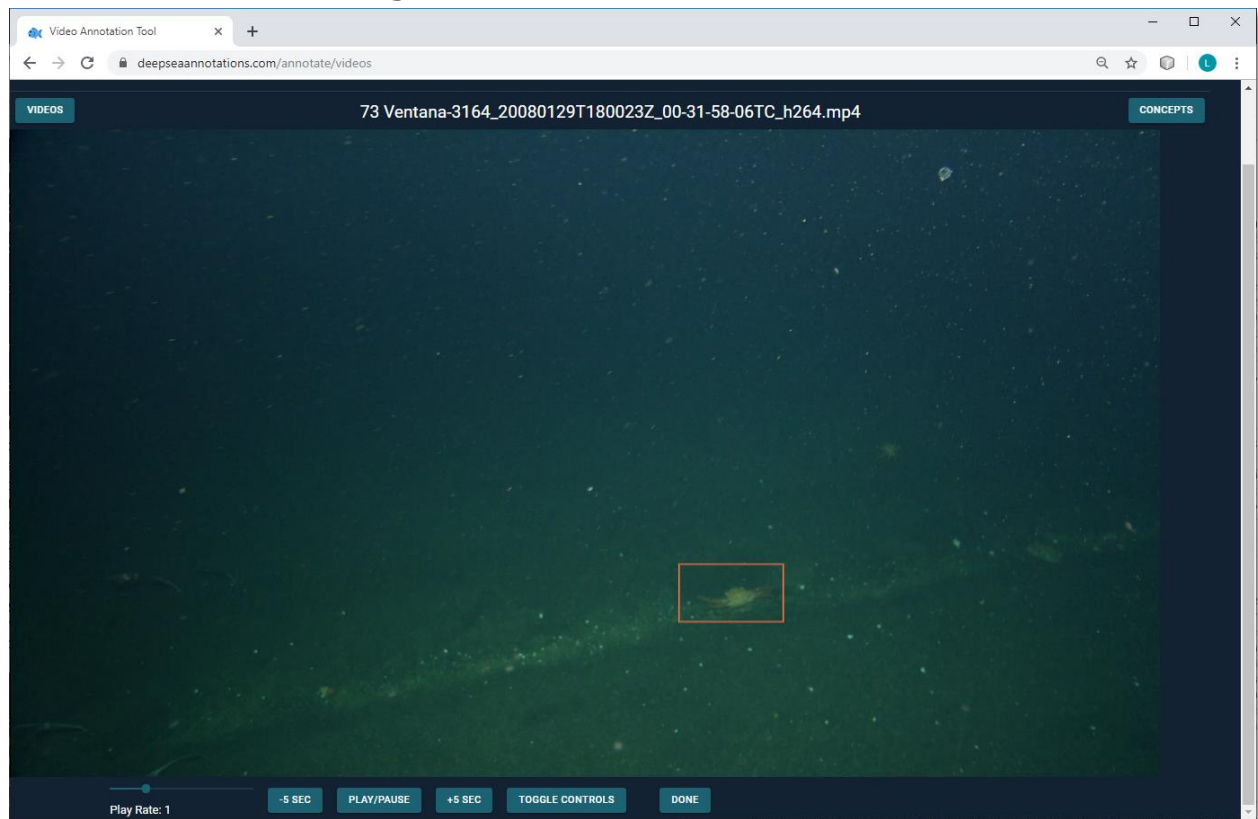
Details about the functionality of the software can be found in Appendix A (User-in-the-Loop Guide). The marine biology student annotators used the website to create more than 40,000 human annotations of more than 100 different concepts and more than 600,000 additional annotations from tracking user annotations in the videos (each tracking can last up to 6 seconds and the video contains 30 frames per second). Figure 1 shows an example screenshot of an annotation.

The top 10 annotated concepts are shown next, together with the number of human annotations.

- *Funiculina*, 9220

- *Strongylocentrotus cf. fragilis*, 9165
- *Asteroidea*, 6464
- *Rathbunaster californicus*, 4699
- *Actiniaria*, 4638
- *Ophiuroidea*, 1491
- *Pleuronctiformes*, 913
- *Umbellula lindahil*, 783
- *Porifera*, 749
- *Sebastolobus*, 705

Figure 1: Screenshot of Annotation



Example bounding box of an annotation.

Source: Cal Poly San Luis Obispo and Monterey Bay Aquarium Research Institute.

Phase 2: Train Neural Network Models

More than 100 different CNN models were trained. The results are shown in Appendix B (CNN Model Results). These results show that some models perform well. For example, one of the models achieved 0.90 and 0.81 F1 scores on (1) *Rathbunaster californicus* (starfish) and (2) *Strongylocentrotus fragilis* (sea urchin), respectively, when models included both user annotations and tracking data as inputs.

In this case, the *F1* score is a measurement of how well the automatic annotation model annotates concepts in new videos on which it has not been trained. It is defined as: $1 = 2 * \frac{P * R}{P + R}$, where *P* is the precision and is defined as $\frac{TP}{TP + FP}$, where *TP* is the number of true

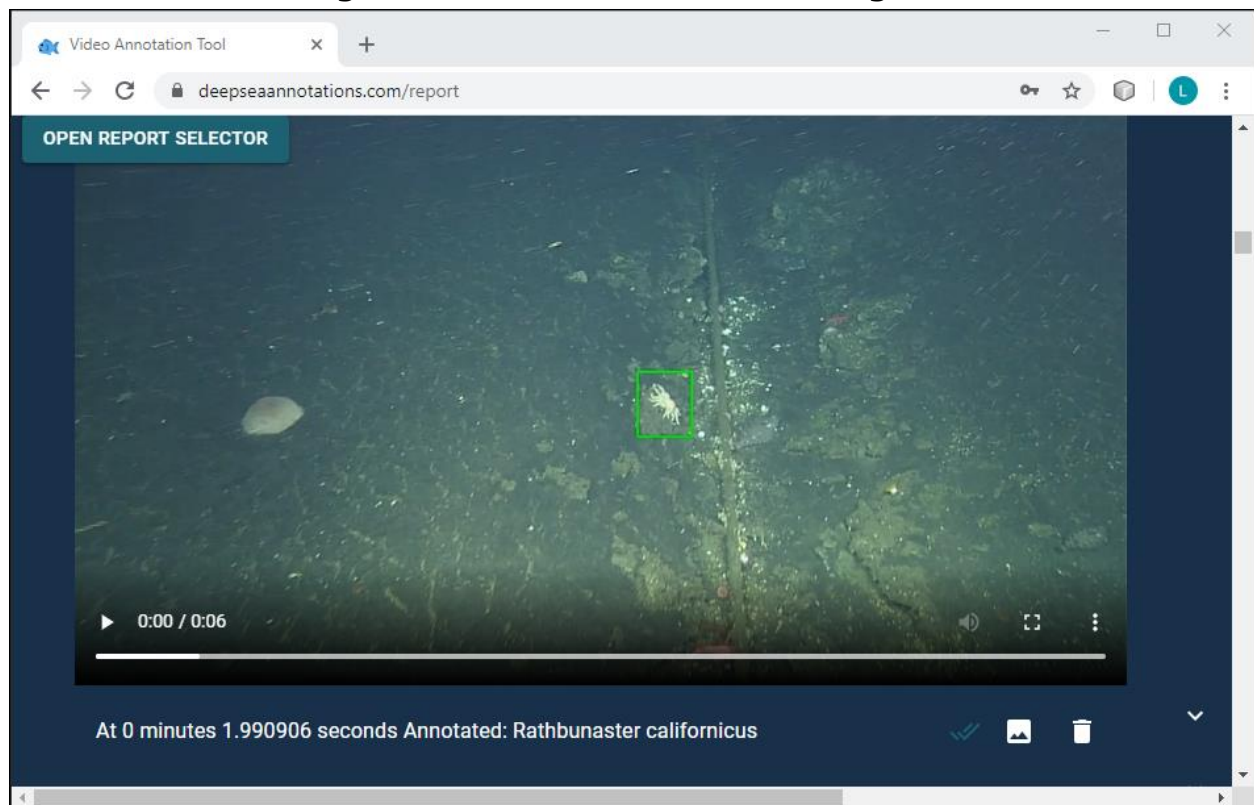
positives — that is, the number of objects that were correctly identified — and FP is the number of false positives — that is, the number of objects that were identified in the video — even though they did not actually exist. Similarly, R is the recall and it is defined as $\frac{TP}{TP+FN}$ where FN is the number of false negatives, that is, the number of objects that were not identified by the model.

The results support the proof of concept that the system can perform well in certain situations. In other situations, the system does not produce good F1 scores; for example, it misses annotations and concepts that are far away and difficult to identify correctly.

Phase 3: Annotation Data Quality Control Tool

The website contains a tool to verify existing annotations. Sometimes, a screenshot alone does not contain enough information to verify an annotation. Therefore, the website also contains the capability to watch a short tracking video around the time the annotation was taken (Figure 2 and Figure 3 contain screenshots of such a video). This enables the annotator to get more information, such as the concept from different angles and the pattern of movement, before deciding whether an annotation is correct. Additionally, an annotator can get more information, such as the depth and location of the video, when verifying an annotation.

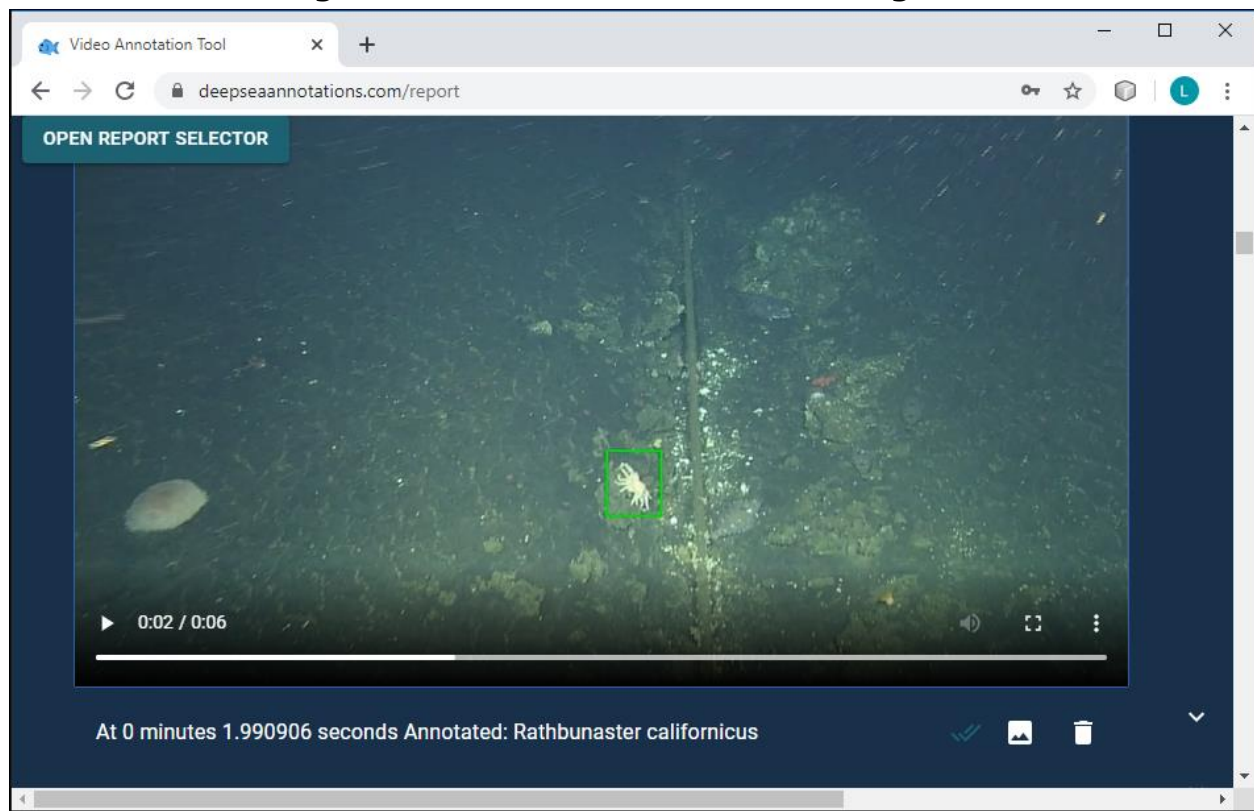
Figure 2: First Screenshot of Moving Crab



Screenshot of the tracking video for a crab (*Rathbunaster californicus*)

Source: Cal Poly San Luis Obispo and Monterey Bay Aquarium Research Institute.

Figure 3: Second Screenshot of Moving Crab



Second screenshot of the tracking video for a crab (*Rathbunaster californicus*)

Source: Cal Poly San Luis Obispo and Monterey Bay Aquarium Research Institute.

The tool enables the user to verify not only the current annotation, but also all other annotations in each annotation frame. This is important because annotated concepts in the annotation frames are used both as positive and negative examples. They are used as positive examples when the annotated concept is one of the concepts that the model is trained on and as a negative example otherwise. The tool also enables the user to verify the quality of the tracking videos. A tracking video that is marked as bad is not used as the source of annotations for training a model.

Lastly, the verification tool is used to verify computer-generated annotations that are created from using a CNN model version to annotate a new video. When a computer-generated annotation is confirmed, it becomes a human annotation. When a computer-generated annotation is modified by a human, the system recognizes that a mistake was made by the current model version. This mistake is recorded in the database and used for reinforcement learning by future model versions.

Phase 4: Automatic Bounding-Box Annotations Displayed in Videos of Organisms






Once a video is annotated by the CNN model, the user can watch the annotated video. The new video is the same as the old video, but a bounding box tracks the occurrence of each object of interest (part of the model) that appears in the video. This video can be used to assess the quality of the model that generated the video, as well as for educational purposes.

Phase 5: Software Displays Video Metadata and Ecological Metrics

A user can click on the information box of a video that is annotated (either by a human or a computer) and see the number of occurrences of each concept, the total number of occurrences, and density information that is calculated based on the total distance covered by the video transect calculated from its GPS start and end coordinates. The distance calculation assumes that the video is taken by an ROV that travels in a straight line. Other information, such as GPS start and end coordinates and depth of video are also displayed.

Figure 4 shows an example of the summary statistics that can be produced for a video.

Figure 4: Screenshot of Summary Table

Summary Table			
Picture	Name	# of Annotations	Creatures per square meter
	Strongylocentrotus cf. fragilis	64	0.471
	Holothuroidea	10	0.074
	Crinoidea	5	0.037
	Asteroidea	6	0.044
	Rathbunaster californicus	32	0.235
TOTAL CONVERT			
total species: 3			
total annotations: 139			
total density(cr/m^2): 1.022			
total distance covered(m): 68			
total depth covered(m): descended 9			

The number of different concepts that are found in a video.

Source: Cal Poly San Luis Obispo and Monterey Bay Aquarium Research Institute.

CHAPTER 4:

Technology/Knowledge/Market Transfer Activities

Current Technology Availability

The source code of the website is available to the general public under the Apache License and can be found at <https://github.com/video-annotation-project>. This includes a detailed guide on how to install and run the software so that users can create their own custom website. The project's GitHub page currently has twelve stars, including from people from MBARI, the Digital Transformation Hub at Cal Poly, Lansrod Technologies, and Nanjing University of Aeronautics and Astronautics. The project has 58 unique cloners.

A webpage describing the project can be found at <https://www.deepseaannotations.com>. The project was featured in Computer Science and Software Engineering News magazine at Cal Poly, which publicized the project. The videos on the website are password protected because they are property of MBARI.

Dissemination of Project Deliverables

The project has been presented twice to the technical advisory committee members. The project and all deliverables have also been shared with MBARI and the Digital Transformation Hub at Cal Poly. The results of the project have also been published at IEEE International Conference on Semantic Computing (Egbert 2020).

Proof of Concept for Future Work

As described in the previous chapters, the technology developed as part of this project is a proof of concept and is not yet ready for large-scale implementation. To date, this project has achieved acceptable accuracy for several species, but a great deal of additional work is needed before this tool can be used for either automated classification of deep-sea organisms or in a workflow for other applications (such as marine or terrestrial birds). Still, the accomplishments of this project have demonstrated that development of a tool to streamline all aspects of workflow for automated classification of species from video is possible. This project has moved all aspects of this workflow to a web-based system.

CHAPTER 5:

Conclusions and Recommendations

The project represents significant advances in annotating underwater video and, ultimately, will facilitate development and implementation of a highly functional system that can generate information critical for monitoring a range of environmental data needed for evaluating the effects of offshore wind facilities and eventual permitting of them.

This project has provided a number of lessons that will be useful for the continued development of this tool. Most importantly, this small project has demonstrated proof of concept that it is possible to transfer the workflow for annotating underwater video to a web-based cloud environment — the tool that can contain all of the steps required to generate and run artificial intelligence models to identify specific concepts/species from video in a single, self-contained platform. This alone offers significant savings in time and expense in annotating underwater video. It also represents a significant advance over existing software that is custom-built and designed for specific applications.

The framework enables users to upload video data; annotate video for specific concepts or targets; conduct quality control on those annotations; generate, evaluate, and improve CNN models that can successfully identify those targets; and ultimately use those models on novel videos. This workflow is capable of generating critical environmental data needed to evaluate the potential effects of energy facilities quickly, efficiently, and cost effectively. Such costs savings should reduce the time and expense required to develop new sources of energy, which will ultimately reduce the levelized cost of energy for California ratepayers. In addition, this framework can be used for a wide range of other applications that require analysis of video or still data, ultimately increasing the amount of data available to evaluate such projects and reducing the costs associated with evaluation and development.

Lessons Learned

In addition, this project provided many other lessons, including the following.

1. The importance of clear communication between annotators and software developers. As annotation proceeded, the annotators were able to continually suggest features that would make their work more efficient. In some cases, these features were as simple as adding keyboard shortcuts, changing the layout of menus, or moving components to different locations on the annotation screen. Each of these features helped increase the efficiency of annotation, which helped reduce the time and costs associated with annotation. The annotation team held online and in-person discussions about these features. The project manager was then able to communicate clear needs to the software development team, again both online and in person. Together, these teams prioritized development of these new features based on urgency and complexity.
2. The need for efficiency in annotation. The process of communication between the annotation and software development teams allowed the project to quickly and efficiently add features to increase the annotation efficiency. Since input data is critical to the development of CNN models, anything that can increase the efficiency with which

humans can annotate videos will reduce the time required to create training data, which equates to lower overall project costs for any project.

3. The number of annotations required. At the start of the project, it was unclear how many annotations would be needed to effectively train the CNN models. Initial runs suggested that several thousand examples of each species would be required, but the development of the tracking feature may have greatly reduced that number. The tracking feature allows the software to track an object forward and backward through time on the video from a single annotation in a single frame. Each of these tracked frames may provide additional training data; with high resolution video (such as 60 frames per second) even an object on a video for only a few seconds may provide 10 – 100 additional annotations for each human-generated annotation. The preliminary results from this project suggest that CNN models may perform well with as few as 1,000 human-generated annotations for a species. The researchers observed that moving from 1,000 to 5,000 annotations improves the F1 score by around 5 percent. Therefore, 1,000 annotations are enough to have a good predictive model, but increasing the number of annotations, particularly verified annotations, can result in better accuracy, perhaps F1 scores above 0.9.
4. Other factors that influence model performance and annotation efficiency. The predictive skill of the models also depends on the complexity of the target as well as its similarity to other potential objects, whether or not they are targets. Targets that are more similar visually to other objects will likely require much more training data for positive (that is, the target) and negative (that is, objects that are not the target) examples. Furthermore, the number of annotations depends on the abundance of the targets in the videos. Targets that are abundant will necessarily require less overall video — and therefore less annotator time — than targets that are rare.
5. The need to be able to efficiently verify and quality check annotations. Initially, the work plan called for human verification of human-generated annotations to occur later in project development. However, it became clear that as soon as CNN model training was initiated, erroneous annotations greatly hindered model performance. Therefore, the software development team prioritized creation of tools to efficiently facilitate human verification of annotations. This verification and quality-assurance/quality-control feature is absolutely essential to ensure that CNN models are trained only on annotations that have been confirmed as correct.
6. The development of collections. As annotation, verification, and CNN model development continued, it became clear that the teams needed a way to subset or select sets of concepts, annotations, and videos. These collections enabled users to identify and select subsets that met a certain criterion. Therefore, the software development team created a feature that enabled users to define specific collections of objects. For example, some of the videos were much higher quality than others — and therefore more effective for both annotation and CNN model evaluation. Using the collection feature, users were able to identify only these high-quality videos, so subsequent annotation, verification, and training would be able to quickly and easily select only these videos. Similarly, users were able to include only those annotations that had been verified from specific species and/or videos to use for training CNN models.

Conclusions and Future Work

This project has provided clear proof of concept that it is possible to develop an efficient and cost-effective web-based platform that can be used to conduct all steps in the workflow required to create CNN models that can generate useful data for any environmental monitoring that uses video data. It focused on identification of deep-sea benthic organisms that could be affected by OSW energy development, but the potential applications for this tool are practically limitless. It could be used to evaluate densities of seabirds or marine mammals around potential OSW locations if video were available from these locations. During development of OSW facilities, permitting might require installation of video cameras on turbines, which could be used to monitor these important species in real time, providing data to assess the environmental effects. Such a system could potentially allow for mitigation measures in real time, such as slowing or halting turbines when vulnerable species are present. Similarly, such a system could easily be adapted to provide similar information for existing or other proposed facilities, such as solar farms or terrestrial turbine installations.

However, to make these applications a reality, some additional work is needed. Much of this work is described in Chapter 4 and includes several steps.

1. Additional investment in the user interface. The user interface is often one of the challenging aspects of any software project because the developers need to anticipate nearly any action that a user may take. In addition, the user interface needs to be highly intuitive, so users can learn to navigate it with as little training as possible. The current user interface was developed in collaboration between the annotators (the primary users) and the team of software developers. The annotators used the software and suggested features or improvements to the software team, who made the changes. The team used other online tools (such as Google sheets and slack) to identify and prioritize features. Priority for these improvements was determined by urgency and level of complexity, with more urgent and less complex features added first. In its current state, the user interface is functional and works well for the small group of users that interface with it on a regular basis, but it is not yet sophisticated enough to anticipate all potential user actions. Fortunately, rich literature exists on improving user interfaces (or 'user experience'). This work will require time and continued collaboration between the annotators and the software developers, but the collaborative model used for this project will work well in any future project.
2. Generation of additional training data. Prior to the project, it was unclear how much training data would be needed for a given species or how long that data would take to create. The project revealed that training a model takes several hours, and models perform better when they are trained on more than 1,000 annotations.
3. Alternative algorithms. Because of the short time frame and limited funding, this project used only one CNN algorithm (RetinaNet). It has performed reasonably well for the few species for which there is sufficient training data, but it has not been possible to evaluate the performance of models using this algorithm compared to others. It is possible that other algorithm architecture may perform more accurately or efficiently, be faster, or require less training data. Future work should consider alternative algorithms.
4. Hierarchical models. In some cases, models may be able to identify a target as a member of a larger group, but not be able to distinguish a specific subgroup. For

example, *Rathbunaster* is a specific species of deep-sea sea star, but at times the computer may not be able to identify it as that species. However, it would still be useful for the system to recognize it as a sea star, even if it cannot determine what species of sea star. This hierarchical workflow is critical, especially for biological organisms that inherently belong to hierarchical groups. Developing this approach, where models recognize that every "*Rathbunaster*" sea star is also a sea star, would be a tremendous advance for any automated classification system that seeks to identify biological entities. This project has made preliminary advances in this area, but more research is required.

5. Novel set of training data. A new application (such as surveying seabirds or benthic species) would require generation of a completely novel set of training data. This process would be somewhat time consuming, but this project has shown that it can be accomplished with a small team of part-time undergraduate annotators supervised by a few experts in a relatively short time frame (approximately 12 months). Presumably, a team of dedicated experts would be able to achieve the same result more quickly, albeit likely at higher cost.

Ultimately, the aim is to develop a fully functional web-based tool that enables users to engage in all steps in this process. This project provides the proof of concept that it can be done and a road map for how to develop a tool that enables users to create and deploy artificial intelligence to identify and enumerate monitoring targets without expertise in computer science. Deploying this tool would lower the cost of generating such information for a range of applications.

CHAPTER 6:

Benefits to Ratepayers

Technological Benefits

To determine how much and where OSW facilities should be developed on California's Outer Continental Shelf, permitting and siting decisions will require vast amounts of new environmental data. This data is valuable to enable developers, the California Energy Commission, the ratepayers, citizens of California, and other stakeholders to evaluate the potential environmental consequences of siting offshore floating wind farms. Given the potential for offshore floating wind farms to generate significant carbon-free electricity, a thorough evaluation of the costs and benefits of this technology in California will be critical to determining if it is one that California should adopt. Data from challenging ocean conditions are typically expensive to collect and interpret. These costs factor into the price to purchase the power, and hence, the price ratepayers will ultimately pay. Tools and methods that can reduce the cost of data acquisition and analysis will help control ratepayer costs.

Underwater video collected on ROVs is a common source of ocean data, but generating information on what it reveals is expensive and time-consuming. Full annotation of videos from a deep-sea research cruise can take several well-trained experts up to two years. That process is not only costly in terms of the labor required, but it would also result in a delay in other steps in the evaluation and permitting process. Such delays may reduce developer interest, and would certainly increase the cost of any project. However, as described in the previous chapters, this project has made tremendous progress toward developing a web-based system that will allow users to automate identification or classification of deep-sea organisms much more rapidly. There is an upfront cost to develop a database of annotations for the species of interest, on the order of \$1,000 per species to achieve reasonable predictive accuracy in interpreting new video. This database can continue to be used for new videos in additional areas with similar marine wildlife communities, so the recurring cost is primarily to run the CNN models on new videos. The time required to run the model is miniscule compared to the time needed to visually interpret an entire video. Moreover, the project demonstrated that the annotations could be produced by trained marine biology undergraduate students whose efforts were verified by an expert. In other words, the expert could focus on viewing frames in which the target species appear and can ignore the large fraction of the video where no marine organisms are visible.

This project has described the subsequent steps that would be required to develop a fully deployable web-based tool to facilitate this analysis. Such a tool would greatly reduce the costs to evaluate any information or video generated from specific sites of interest and has the potential to reduce analysis time from years to potentially as little as weeks, thus greatly reducing the overall costs to generate the essential data. In addition, the platform would be flexible and could be used to evaluate the effects of ongoing operations, potentially identifying locations or activities where effects are greater, enabling project managers to engage in adaptive management. Furthermore, the workflow that this tool uses could be efficiently adapted to generate information for other systems, such as monitoring of effects of offshore or onshore wind facilities on birds or any other process that could require video monitoring of

a set of targets or processes. Widespread use of this tool would streamline and reduce the costs of generating a range of important environmental or scientific information.

Finally, the source code for the tools developed here is freely available for any user to develop applications to further improve functionality and workflow.

LIST OF ACRONYMS

Term	Definition
AI	artificial intelligence
AWS	Amazon Web Services
Cal Poly	California Polytechnic State University San Luis Obispo
CNN	convolutional neural network
MBARI	Monterey Bay Aquarium Research Institute
OSW	offshore wind
ROV	remotely operated vehicle

REFERENCES

- Graabak, I., and M. Korpås. 2016. Variability Characteristics of European Wind and Solar Power Resources—A Review. *Energies* 9:449.
- GWEC. 2018. GWEC Global Wind 2017 Report.
- IRENA. 2016. Innovation Outlook: Offshore Wind.
- 4C Offshore. 2017. Vindeby Offshore Wind Farm.
<https://www.4coffshore.com/windfarms/vindeby-denmark-dk06.html>. Accessed September 11, 2018.
- 4C Offshore. 2018. Hywind Scotland Pilot Park Offshore Wind Farm.
<https://www.4coffshore.com/windfarms/hywind-scotland-pilot-park-united-kingdom-uk76.html>. Accessed September 11, 2018.
- Bailey H., K. L. Brookes and P. M. Thompson. 2014. Assessing environmental impacts of offshore wind farms: lessons learned and recommendations for the future. *Aquatic Biosystems* 10: 8 (2014). <https://doi.org/10.1186/2046-9063-10-8>.
- Egbert, Hanson, Lubomir Stanchev, and Benjamin Ruttenberg, 2020, Automating Deep-Sea Video Annotation Using Machine Learning, IEEE International Conference on Semantic Computing, pages 17-24.
- Furness, R.W., H. M. Wade and E. A. Masde. 2013. Assessing vulnerability of marine bird populations to offshore wind farms. *Journal of Environmental Management* 119: 56-66.
- James, R., and Costa Ros, M. 2015. Floating Offshore Wind: Market and Technology Review. United Kingdom.
- Henkel S., R. Suryan, B. Lagerquist. 2014. Marine Renewable Energy and Environmental Interactions: Baseline Assessments of Seabirds, Marine Mammals, Sea Turtles and Benthic Communities on the Oregon Shelf. In: Shields M., Payne A. (eds) *Marine Renewable Energy Technology and Environmental Interactions. Humanity and the Sea*. Springer, Dordrecht.
- Lin T., P. Goyal, R. Girshick, K. He, P. Dollar, 2017. Focal Loss for Dense Object Detection.
<https://arxiv.org/pdf/1708.02002.pdf>
- Musial, W., and B. Ram. 2010. Large-Scale Offshore Wind Power in the United States.
- Wang, Y., R. Walter, C. White, H. Farr, and B. Ruttenberg. 2019. Assessment of surface wind datasets for estimating offshore wind energy along the Central California Coast. *Renewable Energy* 133, 343-353, doi:10.1016/j.renene.2018.10.008.
- Winds, T., 2016. Unsolicited Application for an Outer Continental Shelf Renewable Energy Commercial Lease Under 30 CFR 585.230.

APPENDIX A:

User-in-the-Loop Guide

This appendix describes how to use the website www.deepseaannotations.com. Specifically, it describes how to upload videos, annotate them, verify the annotations, and use the annotations to train neural-network models that can be used to automatically annotate unwatched videos. This appendix also contains guidelines of how to measure the quality of existing models and ways of improving the accuracy.

Website Documentation

Uploading the Videos

All videos should initially be placed in an Amazon Web Services (AWS) S3 bucket. The location of the S3 buckets should be set in the project environment, where more information about setting up the project environment can be found at the GitHub repository: github.com/video-annotation-project.

Creating a Concept Bag

The first task for every user is to create a concept bag. This is the set of concepts that the user will use. This is done through the `CONCEPTS` tab. Note that additional concepts can also be added to the concept bag on the fly in the other tabs.

There are three different ways of adding concepts to the concept bag in the `CONCEPTS` tab. The first way is to directly type the name of the concept in the `Concept name` text box and press Enter. The concept location in the hierarchy will appear and the user can select the checkbox on the right to select the concept. Figure A-1 shows an example of selecting *Funiculina*, which is the scientific name for sea pan.

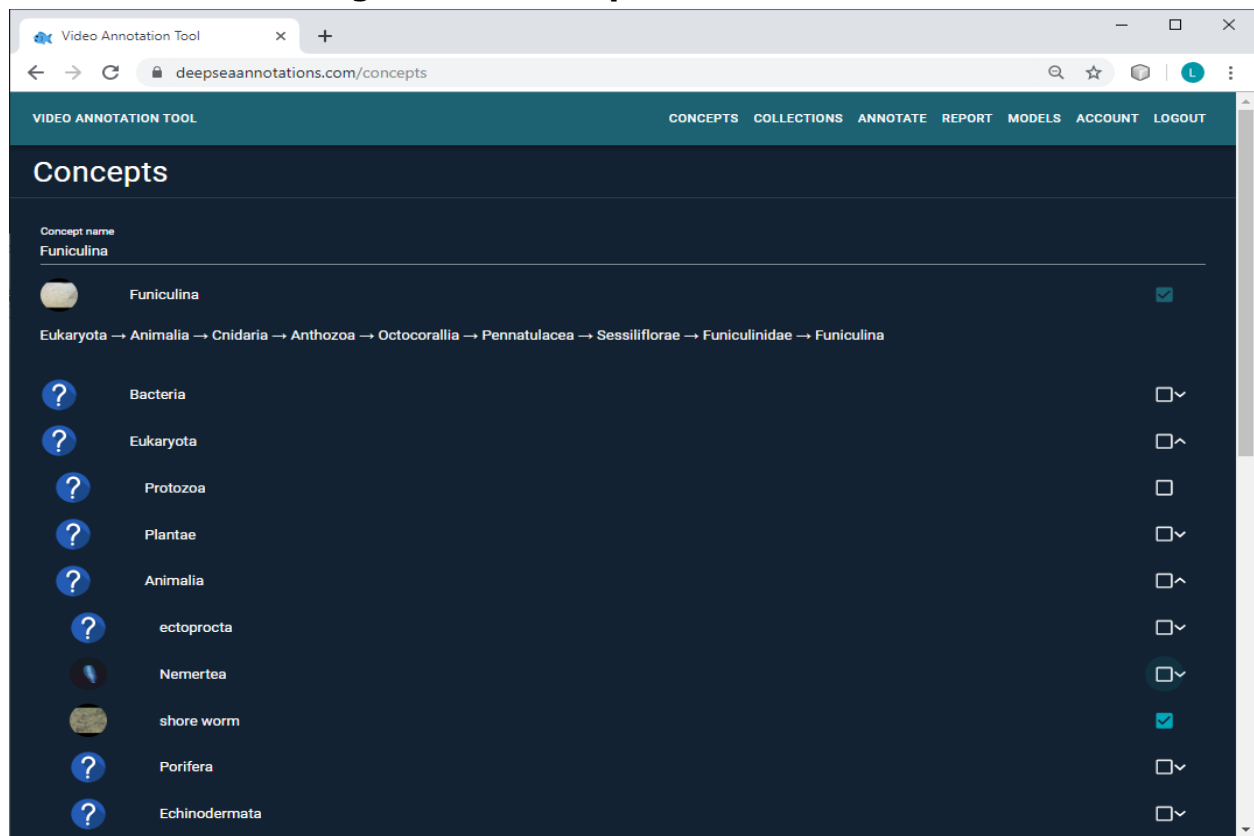
Note that the website supports autocomplete and possible suggestions are automatically shown.

The second way of adding a concept to the concept bag is by expanding the taxonomic tree at the bottom of the screen. This can be done using the up and down arrows on the right. Once the desired concept is found, the checkbox on the right side needs to be selected. Note that preloaded images are associated with some of the concepts.

The third way of adding a concept to concept hierarchy is by typing the concept number in the `Concept name` textbox.

The concept hierarchy and the images that are associated with the concepts are preloaded on the S3 bucket in a `JSON` file and a set of `.jpg` files, where more information on setting up these files can be found at the install documentation in the `github` repo.

Figure A-1: Concept Selection Interface



Shows how to select a concept for a user's concept bag.

Source: Cal Poly San Luis Obispo and MBARI

Creating Collections

There are three types of collections that can be created from the `COLLECTIONS` tab: Annotations, Concepts, and Videos.

In the `Annotations Collections` menu, one must first select an existing collection from a drop-down menu of annotation collections or create a new one. To create a new annotation collection, just click on the `NEW ANNOTATION COLLECTION` button. This is a two-step process, where the system asks for the collection name and its description. Alternatively, the user can select an annotation collection and click the `DELETE THIS COLLECTION` button to delete it.

The interface provides a `COLLECTION INFO` button that gives more information about an annotation collection. These include the concepts in the annotations, the number of annotations for each concept broken by type, the users that made the annotations, and the videos from which the annotations came. Figure A-2 shows an example of a `COLLECTION INFO` window.

Figure A-2: Annotation COLLECTION INFO window

4 Concepts					
Concept	User Annotations	Tracking Annotations	Verified User Annotations	Verified Tracking Annotations	All Annotations
Actiniaria	1	0	1	0	1
Anoplopoma fimbria	483	41309	409	0	41792
Asteroidea	5	0	5	0	5
Chionoecetes tanneri	444	40842	318	38	41286
Funiculina	2	0	2	0	2
Pleuronectiformes	2	0	2	0	2
Rathbunaster californicus	4055	246538	1806	4	250593
Sebastolobus	6	0	6	0	6
splitnose rockfish	1	0	1	0	1
Strongylocentrotus cf. fragilis	8285	203366	2575	0	211651
Synodus	1	0	1	0	1
	0	0	0	0	39115
TOTAL	13285	532055	5126	42	584455
Users (6): adminAli, ali, allison_m, emma, kellen, sierra_s					
Videos (97): 1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 64, 65, 66, 67, 70, 76, 77, 78, 79, 81, 83, 84, 86, 87, 88, 89, 91, 96, 97, 98, 99, 102, 106, 107, 108, 109, 110, 116, 117, 129, 130, 131, 133, 134, 157, 158, 159, 160, 161, 162, 163					

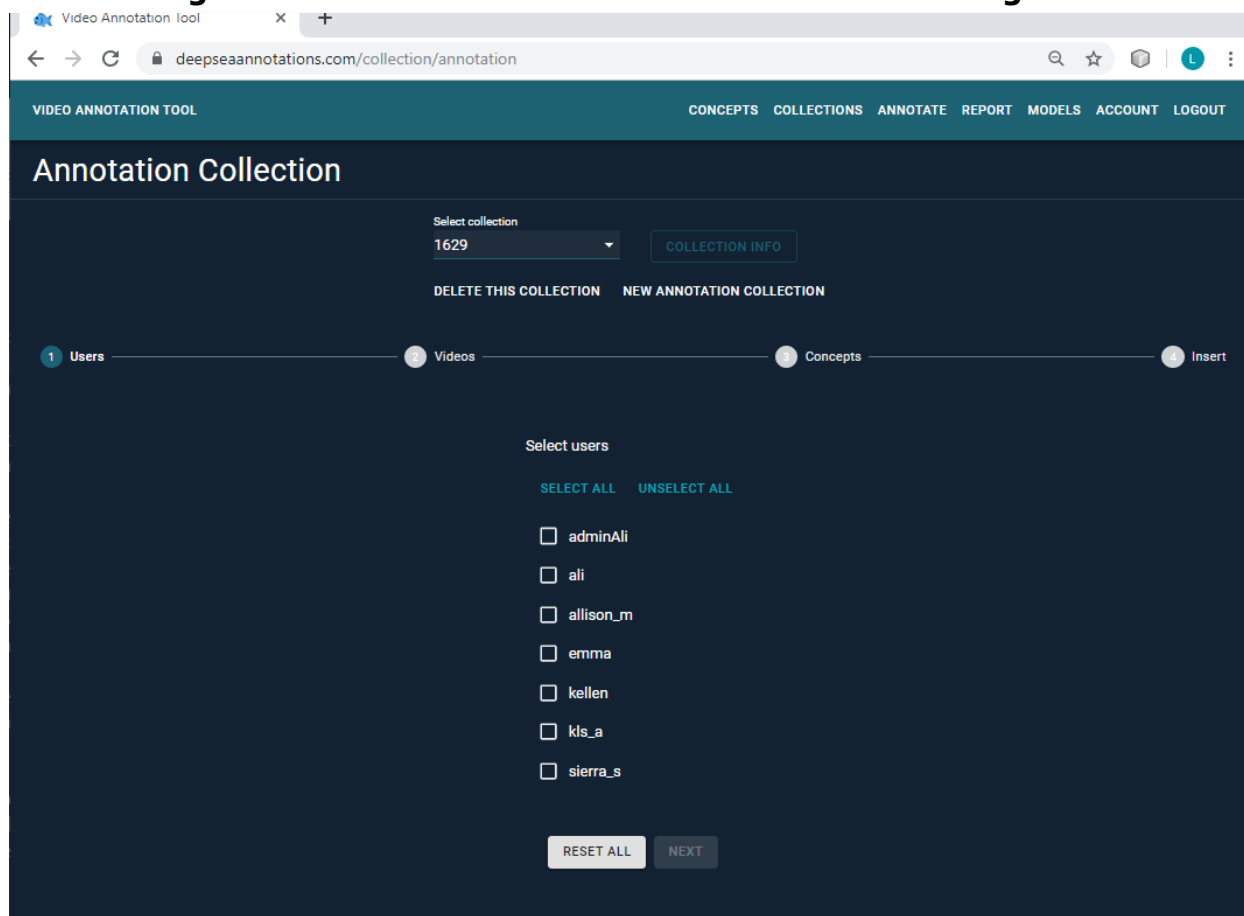
Shows an example description of an annotation collection.

Source: Cal Poly San Luis Obispo and MBARI

Once the collection name is selected, there is a four-step process to add annotations to the annotation collection. Note that for new annotation collection, the process describes the annotations of the collections. However, for an existing annotation collection, the process describes how to add additional annotations to the collection.

The first step is to select the users. Annotations only from these users will be considered. Note the `SELECT ALL` button on the top, which allows us to select all users, and the `UNSELECT ALL` button, which unselects all checkboxes -- see screenshot in Figure A-3.

Figure A-3: Annotation Collection Interface: Selecting Users

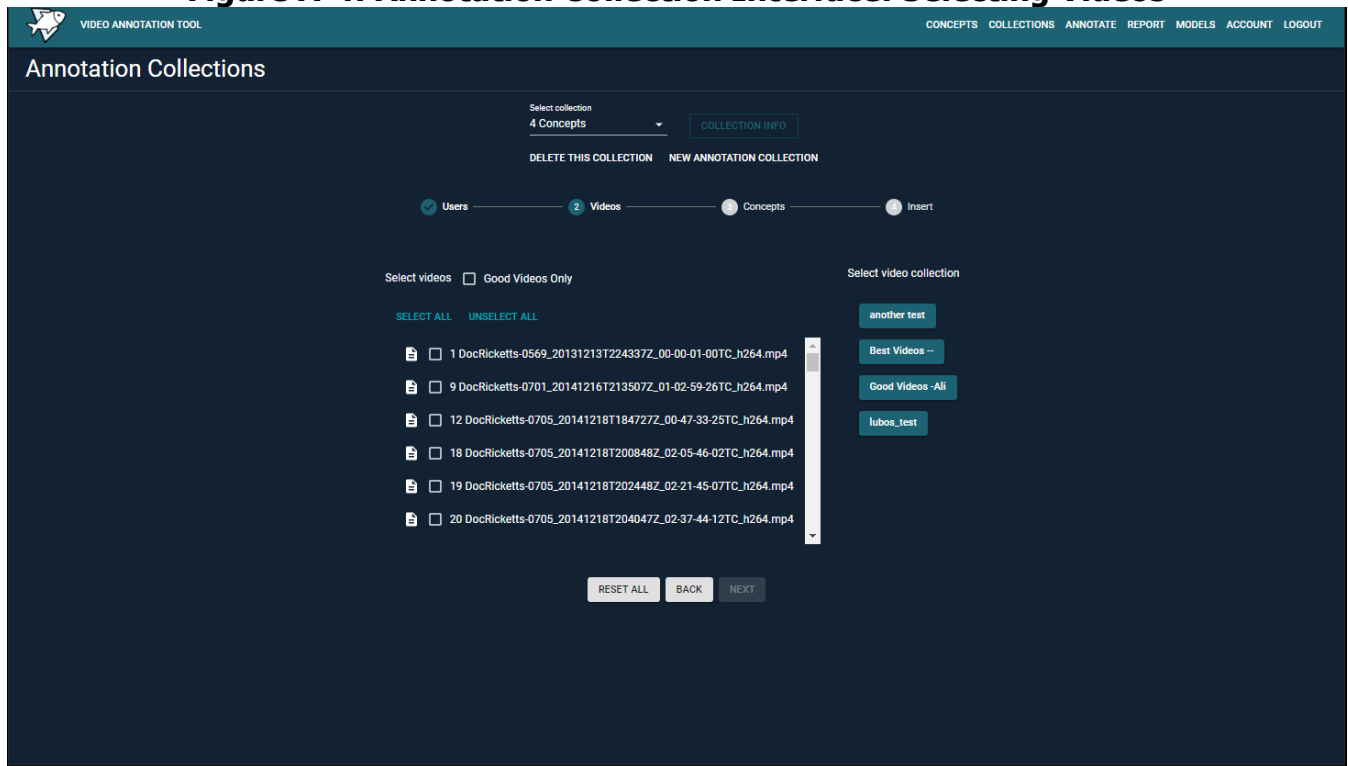


Shows how to select the users when creating an annotation collection.

Source: Cal Poly San Luis Obispo and MBARI

The next step is to select the videos for which one wants to select annotations.

Figure A-4: Annotation Collection Interface: Selecting Videos



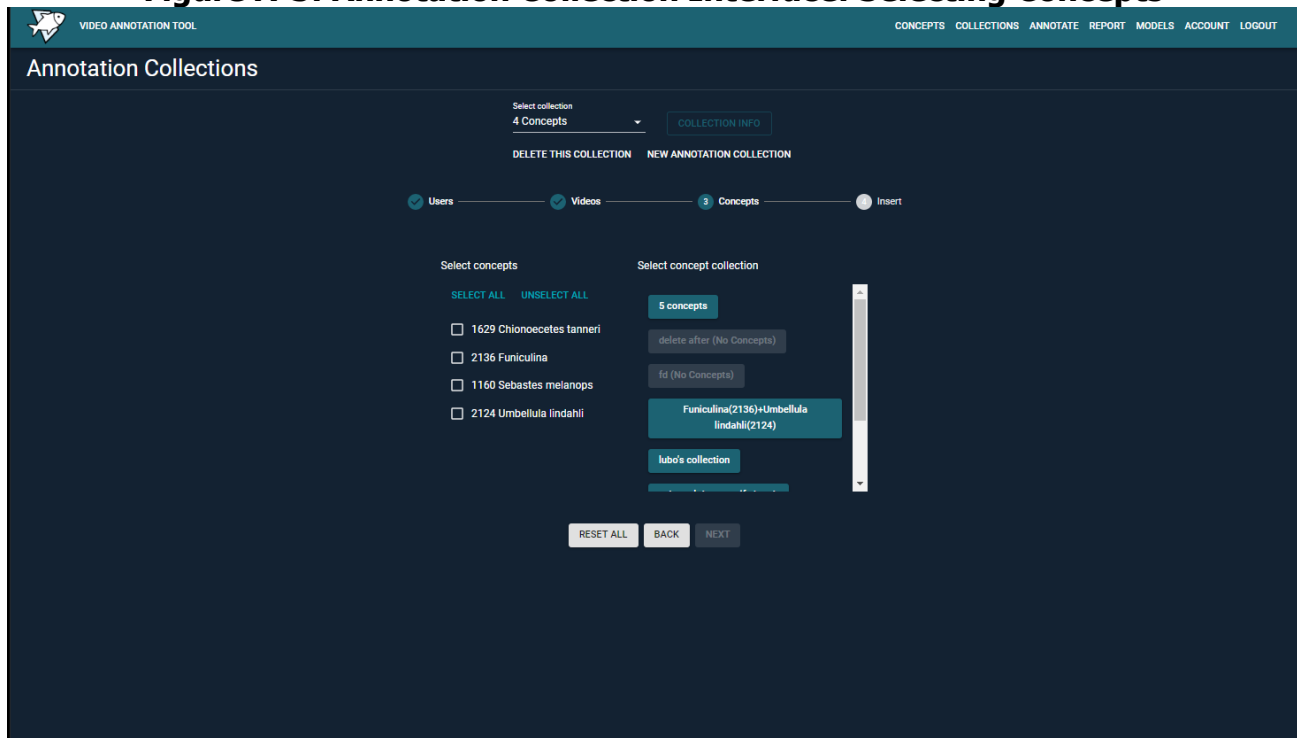
Shows how to select the videos when creating an annotation collection.

Source: Cal Poly San Luis Obispo and MBARI

As shown in Figure A-4, on the left-hand side one can directly select the videos of interest. Alternatively, from the right-hand side one can select a video collection. Once a video collection from the right-hand side is selected, all the videos in the collection appear on the left-hand side with checkboxes. Additionally, a `Good Videos Only` checkbox is provided that allows the user to select only good videos for the annotation collection.

The next step is to select the concepts for which one wants to extract annotations. Similarly to the previous step, the user can directly select the concepts on the left-hand side or select a concept collection on the right-hand side. Note that only the concepts that appear in the selected videos and have been annotated by the selected users are displayed – see Figure A-5.

Figure A-5: Annotation Collection Interface: Selecting Concepts



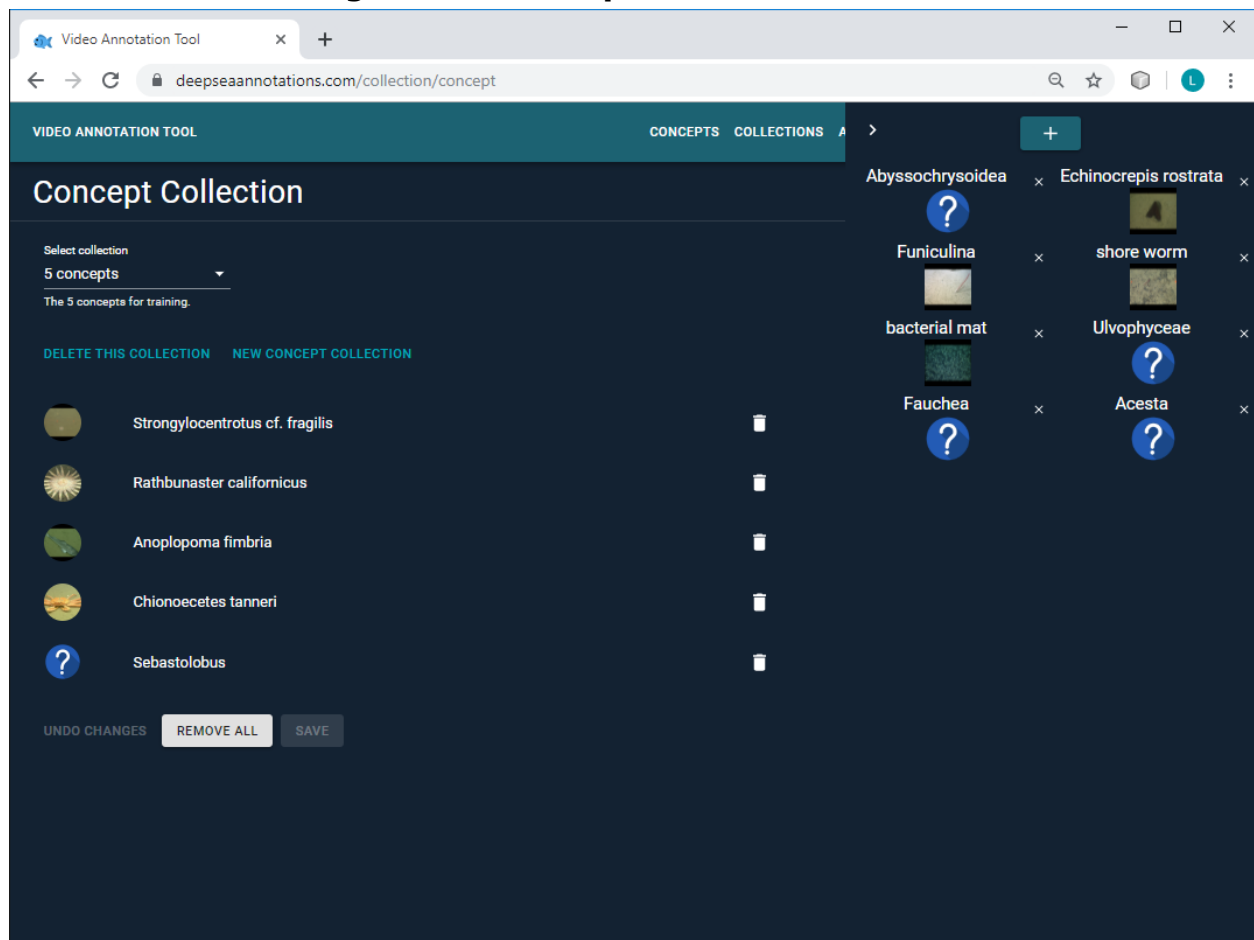
Shows how to select the concepts when creating an annotation collection.

Source: Cal Poly San Luis Obispo and MBARI

The last step shows the number of annotations in the created collection with and without tracking annotations. The user can then create/augment the collection of annotations with or without the tracking data and potentially add more annotations to the collection.

The `Concepts Collections` tab allows one to create a new concept collection or modify an existing concept collection. The available concepts in the concept bag are shown in drawer on the right-hand side that opens when one presses the `CONCEPTS` button. When adding a new concept to a concept collection, the user can select a concept from their bag of concepts or add a new concept that is not in the bag by pressing the "+" button – see Figure A-6.

Figure A-6: Concept Collection Interface



Shows how to select the concepts when creating a concept collection.

Source: Cal Poly San Luis Obispo and MBARI

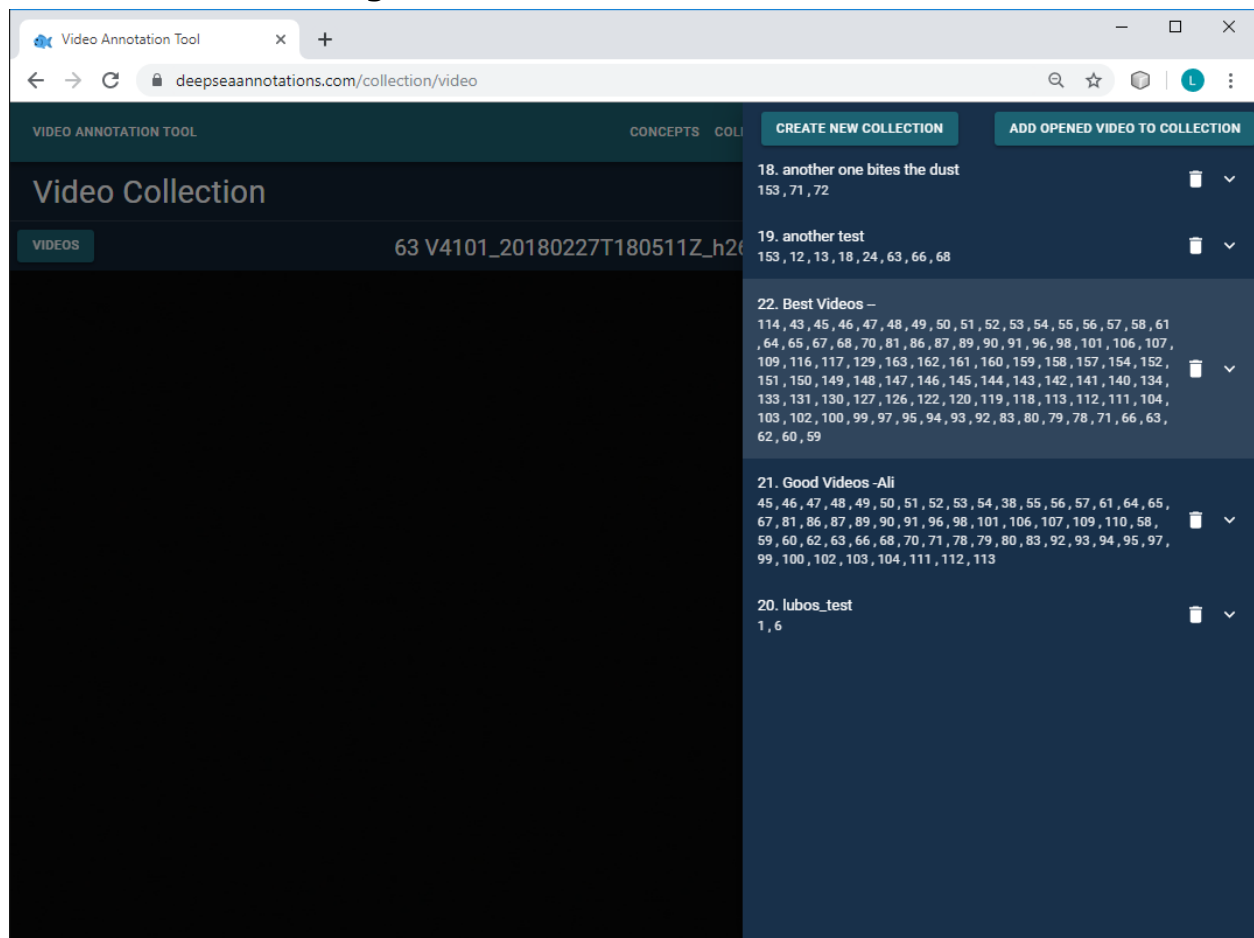
The first step in creating a concept collection is selecting an existing collection or creating a new one by clicking the `NEW CONCEPT COLLECTION` button.

There is also an option to delete an existing concept collection by selecting a concept collection and clicking on the `DELETE THIS COLLECTION` button.

Once a concept collection is selected (or a new empty one is created), one can use the `CONCEPTS` drawer to add additional concepts to the collection. There are three buttons on the bottom of the screen. The `UNDO CHANGES` button undoes the last change, the `REMOVE ALL` button clears the concept collection, while the `SAVE` button makes the changes permanent. There is also a remove icon next to each concept in the collection. If clicked, then the concept is removed from the concept collection.

Lastly, the Videos Collections tab allows the user to create a collection of videos. The menu has two drawers. The right drawer is the Video Collections drawer that displays the current set of video collections – see Figure A-7.

Figure A-7: Video Collection Interface



Shows how to examine the different video collections.

Source: Cal Poly San Luis Obispo and MBARI

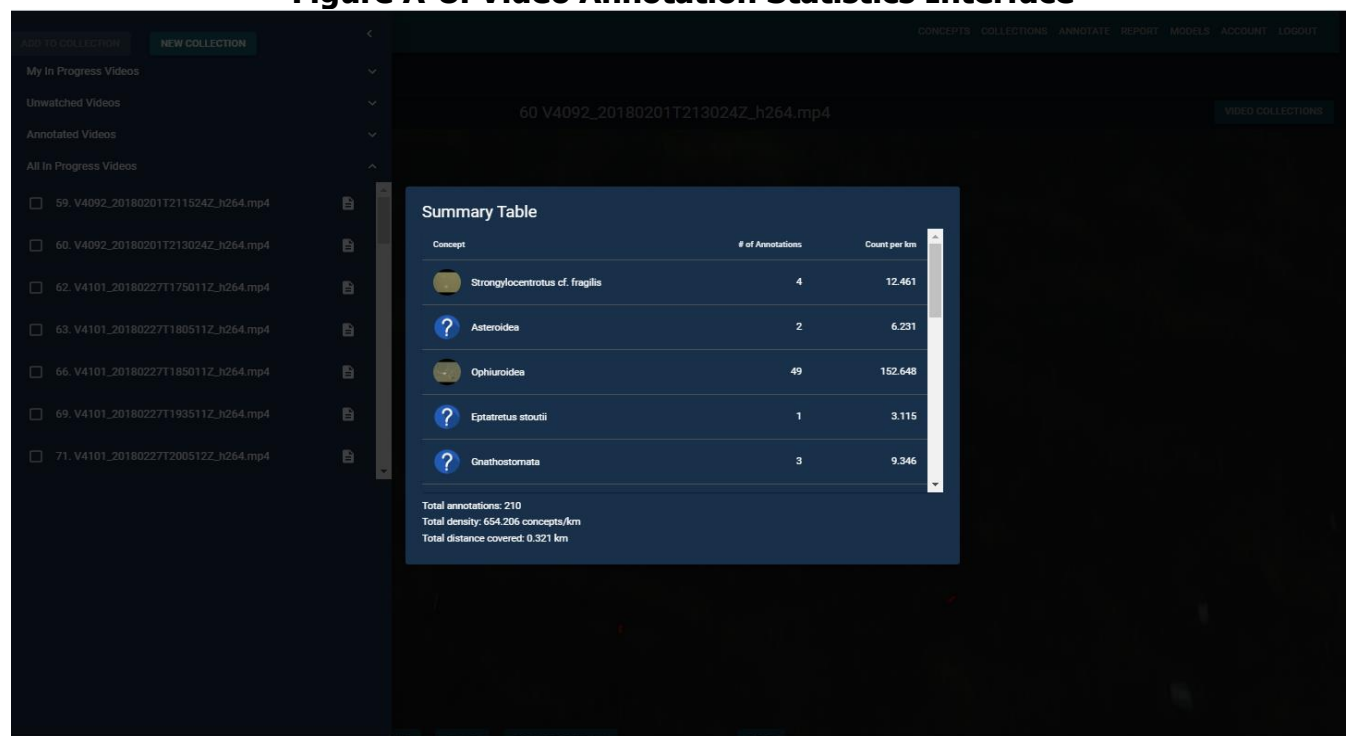
This drawer allows the user to add the current video to an existing collection, create a new collection, or delete an existing collection. The left drawer allows the user to play videos in order to get familiar with their content. The software supports four lists of videos: *My In Progress Videos*, which keeps track of the videos that are currently annotated by the user, *Unwatched videos*, which contain no annotations, *Annotated Videos*, which have been fully annotated by someone, and *All in progress videos*, which are currently being annotated.

Note that there is an info image next to each video. If clicked, then information about the video appears. This includes the name of the video, the users that annotated it, the start and stop GPS coordinates of the video, the start and end depth in meters, the start and end time of the video recording. There is also a video description associated with each video and checkbox to denote whether the video is good or not. Each video can also be tagged as unwatched, annotated, or in progress.

There are two buttons on the bottom of the information box. The *SUMMARY* button opens a new window that contains the species that are identified in the video, their count, and the density – see Figure A-8. The bottom of the popup window shows summary statistics: total human annotations, total density, total distance traveled by the ROV that shot the video. The *UPDATE* button allows users to change the meta data for the video, that is add a description,

make it a good or not good video, and mark it as unwatched, annotated, or in progress – see Figure A-9.

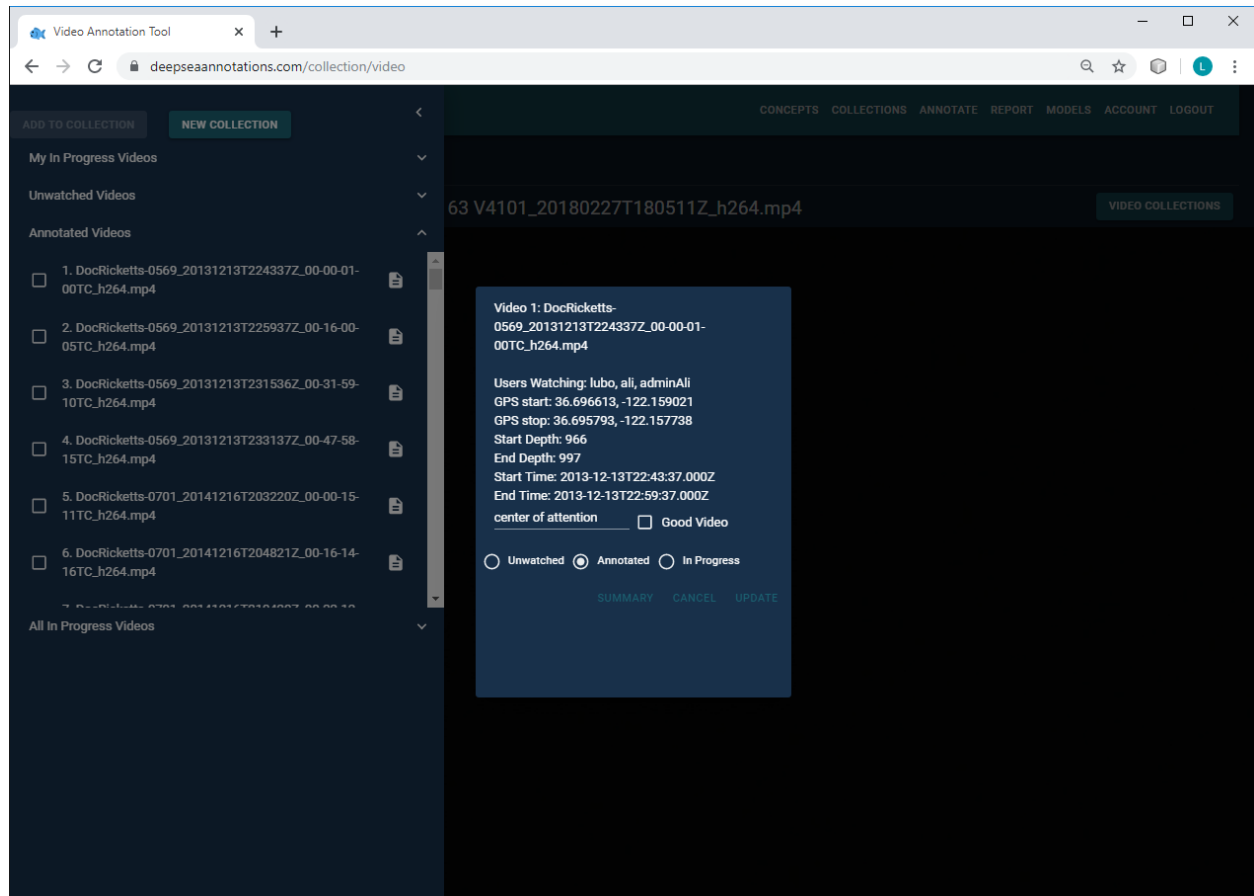
Figure A-8: Video Annotation Statistics Interface



Shows example annotation statistics for a video.

Source: Cal Poly San Luis Obispo and MBARI

Figure A-9: Video Information Interface



Shows an example of how to mark a video as **Good** or as **Annotated**.

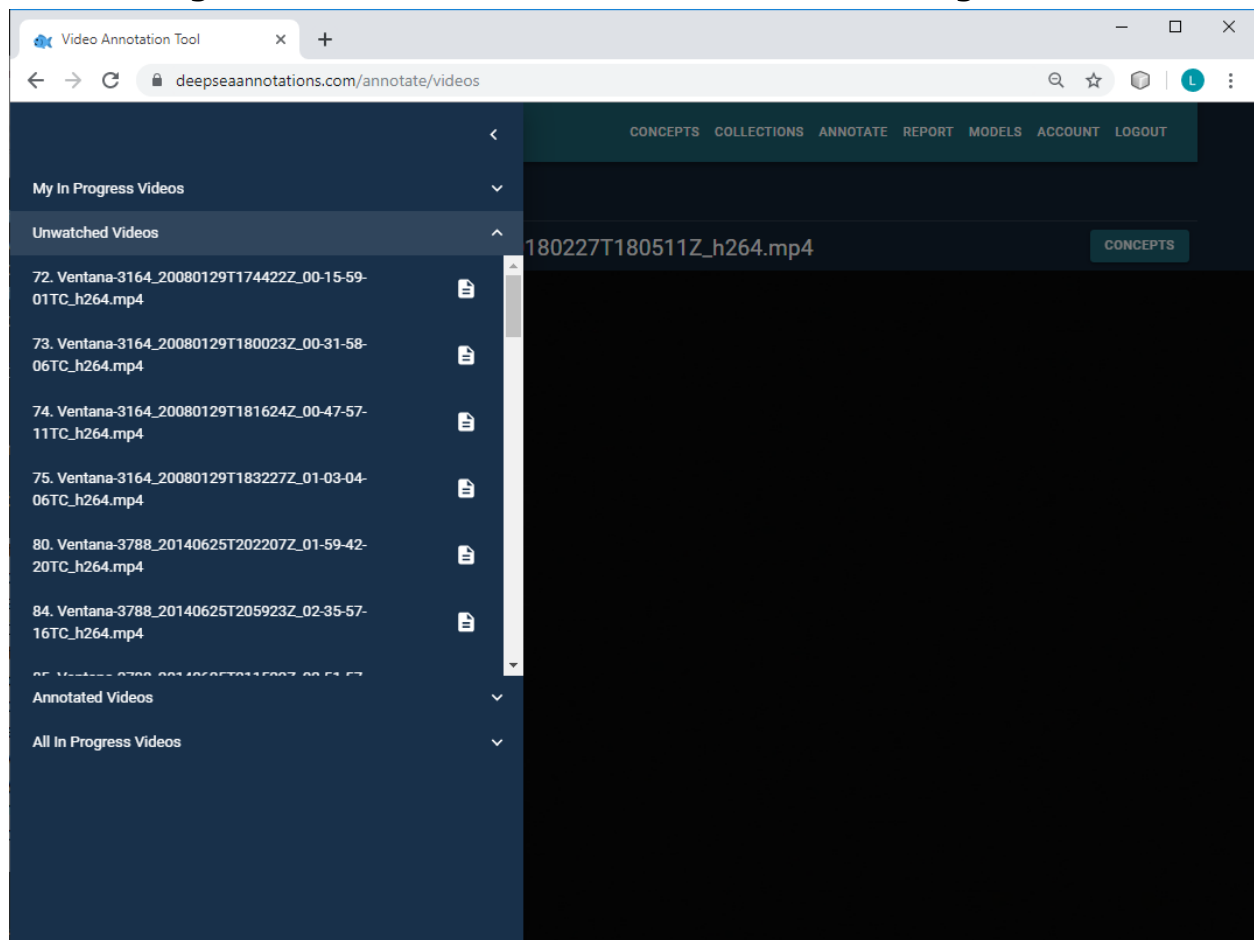
Source: Cal Poly San Luis Obispo and MBARI

Annotate Tab

The **ANNOTATE** tab has two menus: **Videos** and **Verify**.

The **Videos** menu plays a video. First, the user needs to select the video to play from the **VIDEOS** drawer. The user can pick from **My In Progress Videos**, which keeps track of the videos that are currently annotated by the user, **Unwatched videos**, which contain no annotations, **Annotated Videos**, which have been fully annotated and **All in Progress Videos**, which are currently being annotated – see Figure A-10.

Figure A-10: Video Annotation Interface: Selecting a Videos

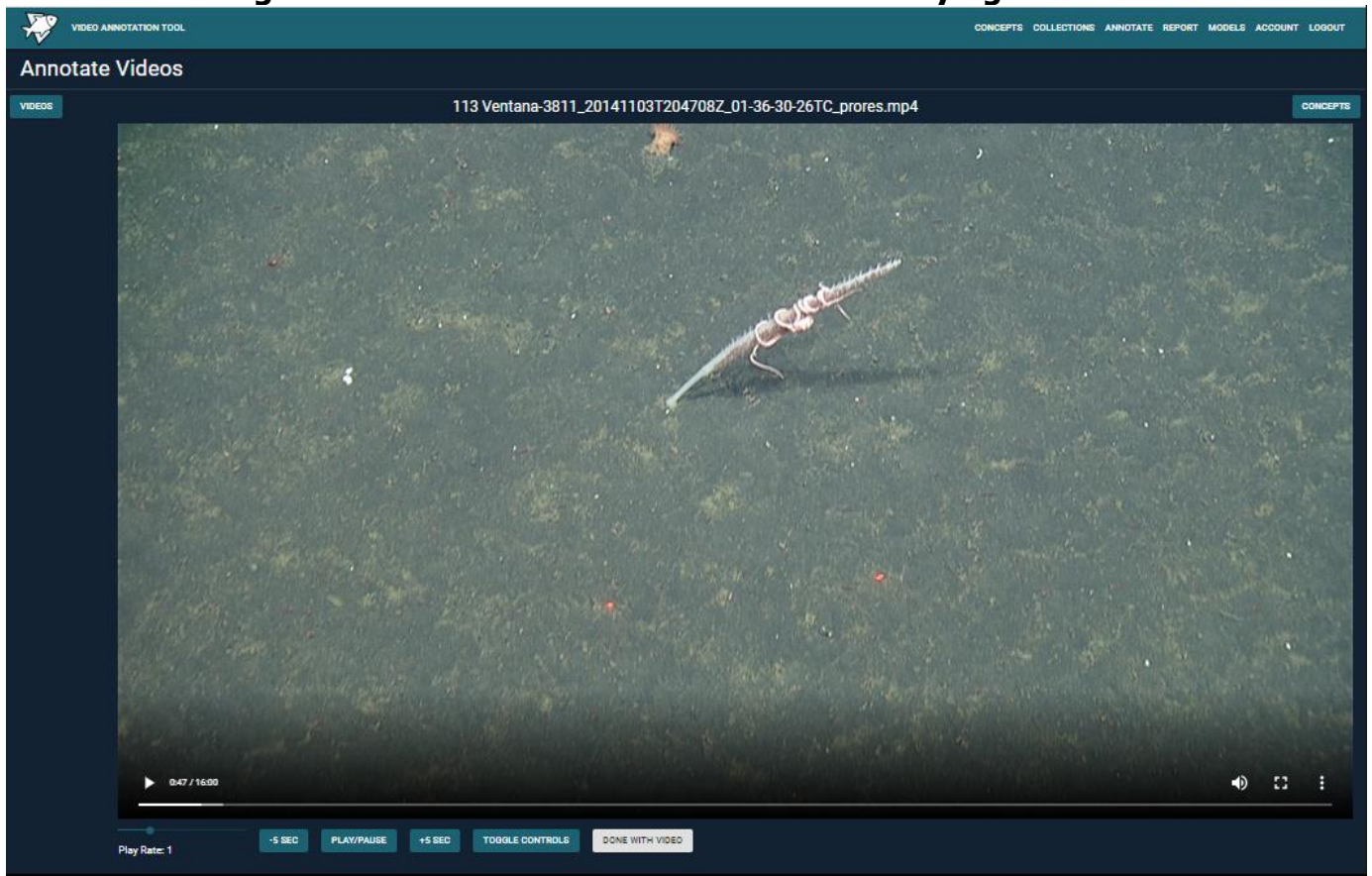


Shows how to select a video to watch in the annotation interface.

Source: Cal Poly San Luis Obispo and MBARI

Next the user can play or pause the video using the `PLAY/PAUSE` button at the bottom of the screen. There are also options to set the play speed, move the video forward or backward by 5 seconds, or toggle the controls. Note that the left and right keyboard keys can also be used to go one second backward or forward and the space keyboard key can be used to play/stop the video. The `TOGGLE CONTROLS` button adds or removes the bar at the bottom of the video that shows the elapsed time and allows us to make the video full screen, adjust the volume, or download the video – see Figure A-11.

Figure A-11: Video Annotation Interface: Playing Videos



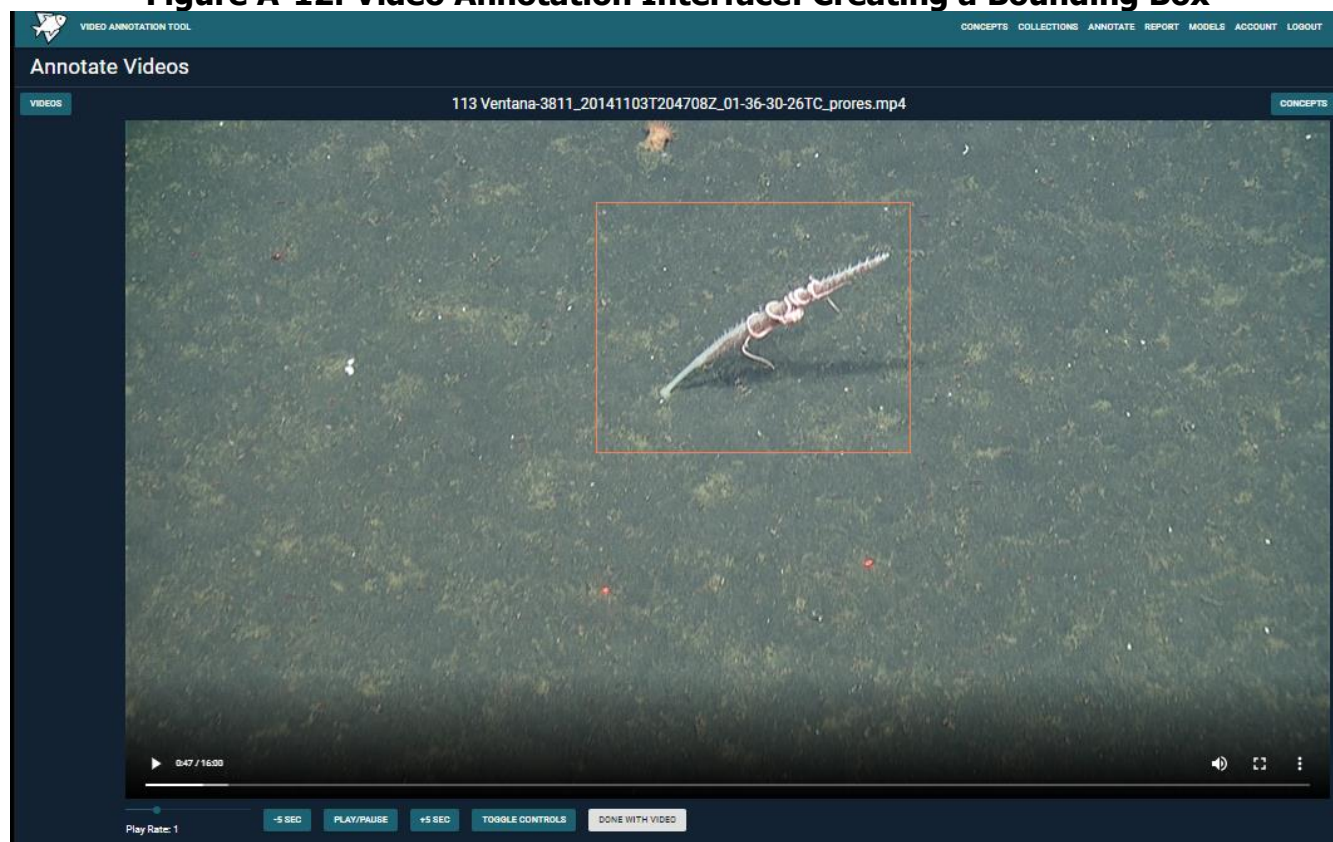
Shows how to play a video using the annotation interface.

Source: Cal Poly San Luis Obispo and MBARI

There is also a `DONE WITH VIDEO` button at the bottom that marks the video as fully annotated.

Once the annotator finds an interesting species to annotate, they can stop the video and create a bounding box around the species – see Figure A-12.

Figure A-12: Video Annotation Interface: Creating a Bounding Box

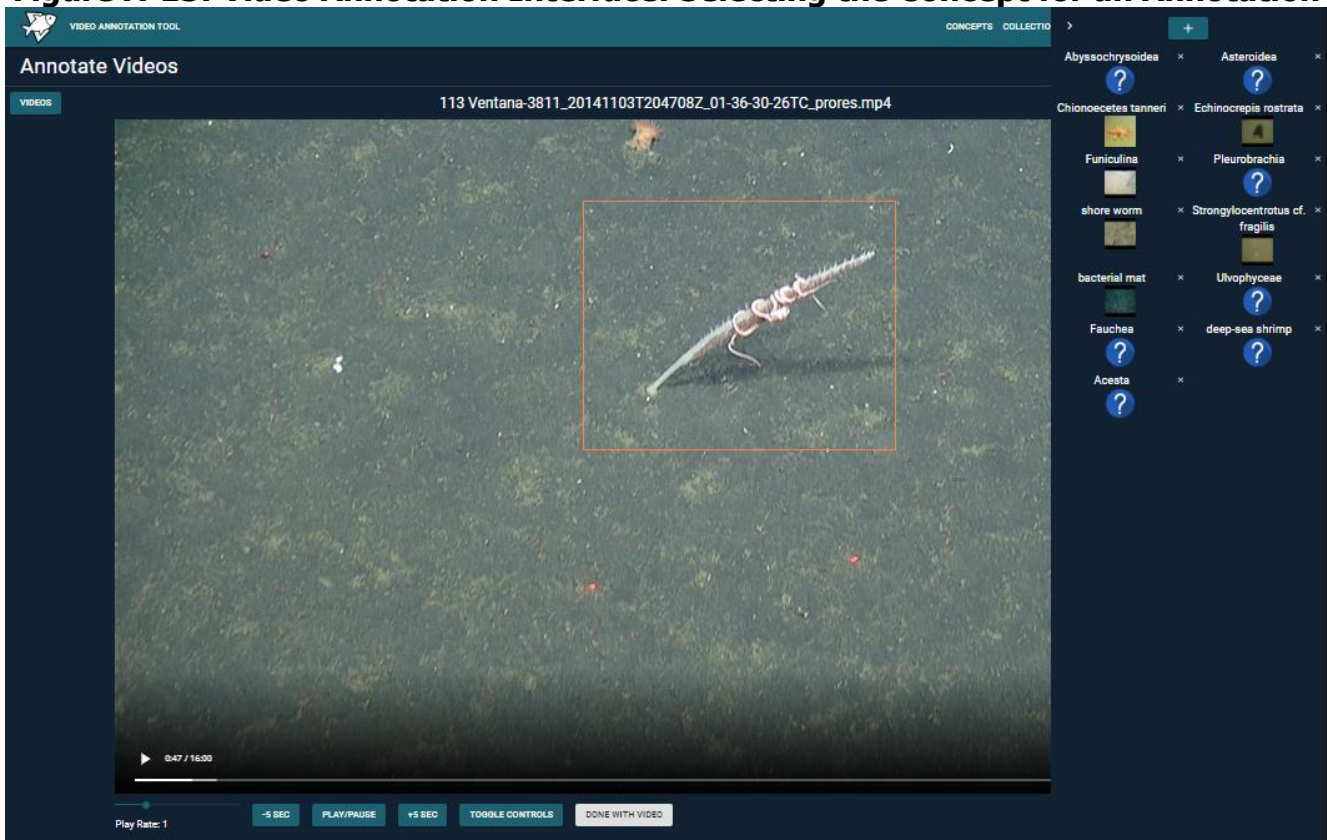


Shows how to create a bounding box using the annotation interface.

Source: Cal Poly San Luis Obispo and MBARI

Next, the annotator can select the concept that is being annotated from the CONCEPTS drawer. The drawer displays the current concepts in the concept bag. However, the user can also press the "+" button to add a new concept to the concept bag – see Figure A-13.

Figure A-13: Video Annotation Interface: Selecting the Concept for an Annotation



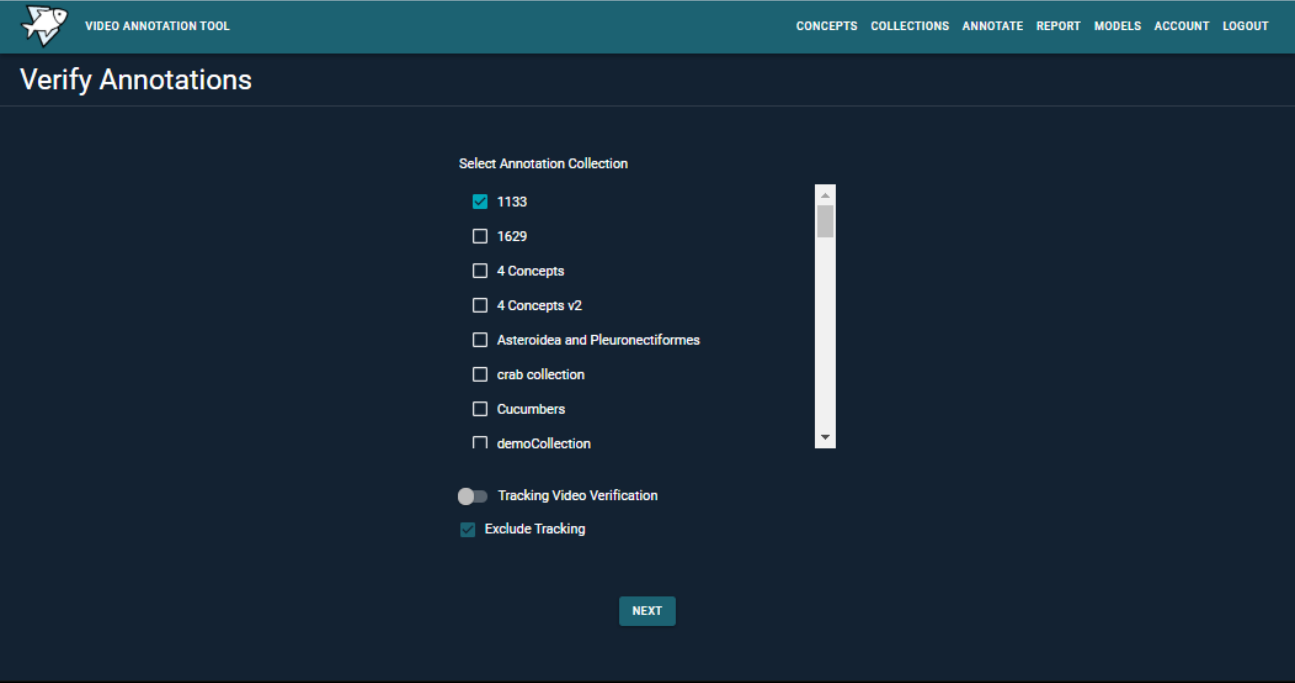
Shows how to selection the annotation concept using the annotation interface.

Source: Cal Poly San Luis Obispo and MBARI

After clicking the concept to annotate, a popup window appears. It allows to add a description of the annotation and possibly mark the annotation it as unsure.

The `ANNOTATE Verify` menu is used to verify an existing collection of annotations. One needs to first select the annotation collection to verify, where it is possible to select multiple annotation collections. Next, one needs to select whether to include annotations from tracking and whether to verify the tracking video for each annotation. By default, tracking annotations are included. If the user chooses to exclude them, then they need to decide whether to enable tracking video verification – see Figure A-14.

Figure A-14: Annotation Verification Interface: Selecting the Annotation Collection to Verify

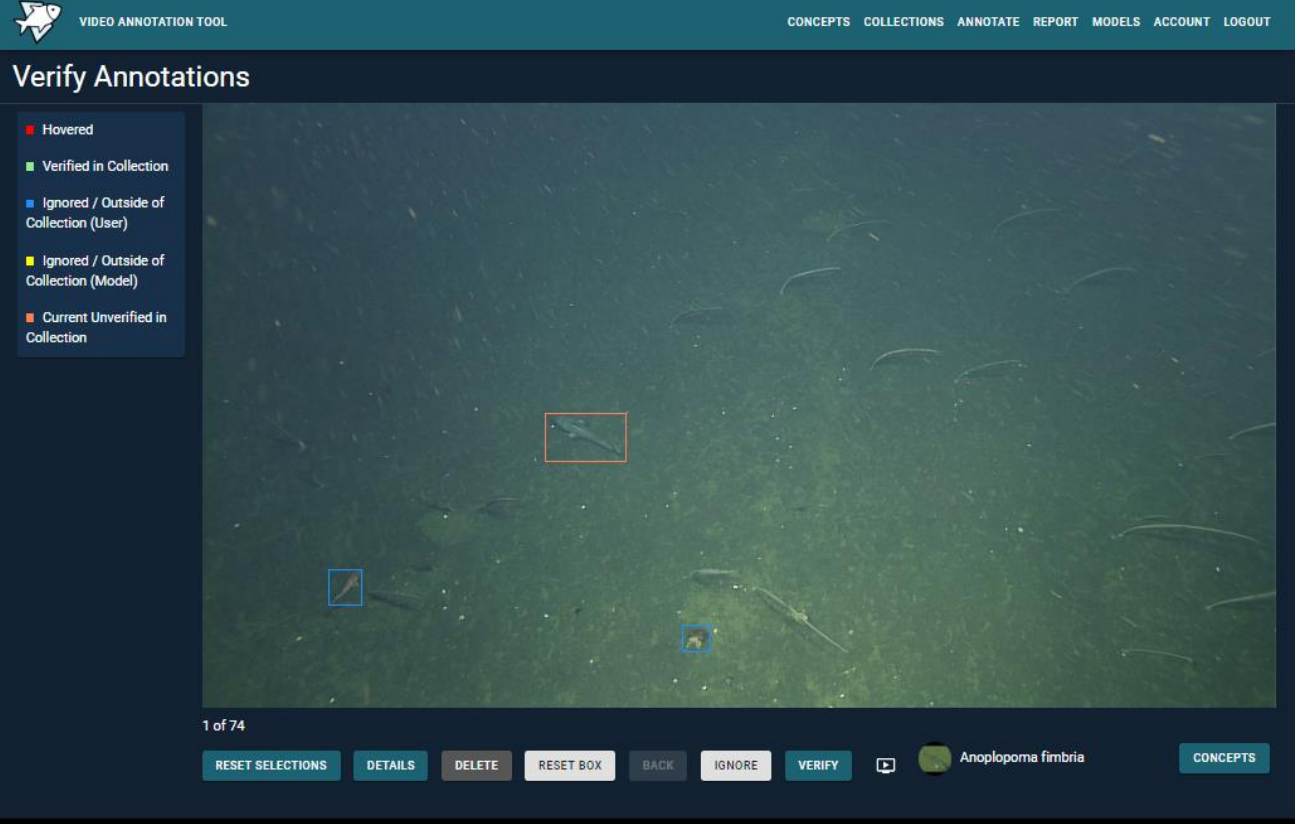


Shows how to select an annotation collection to verify in the Verification Interface.

Source: Cal Poly San Luis Obispo and MBARI

As Figure A-15 shows, five different colors are used for annotation bounding boxes.

Figure A-15: Annotation Verification Interface: Verifying an Annotation

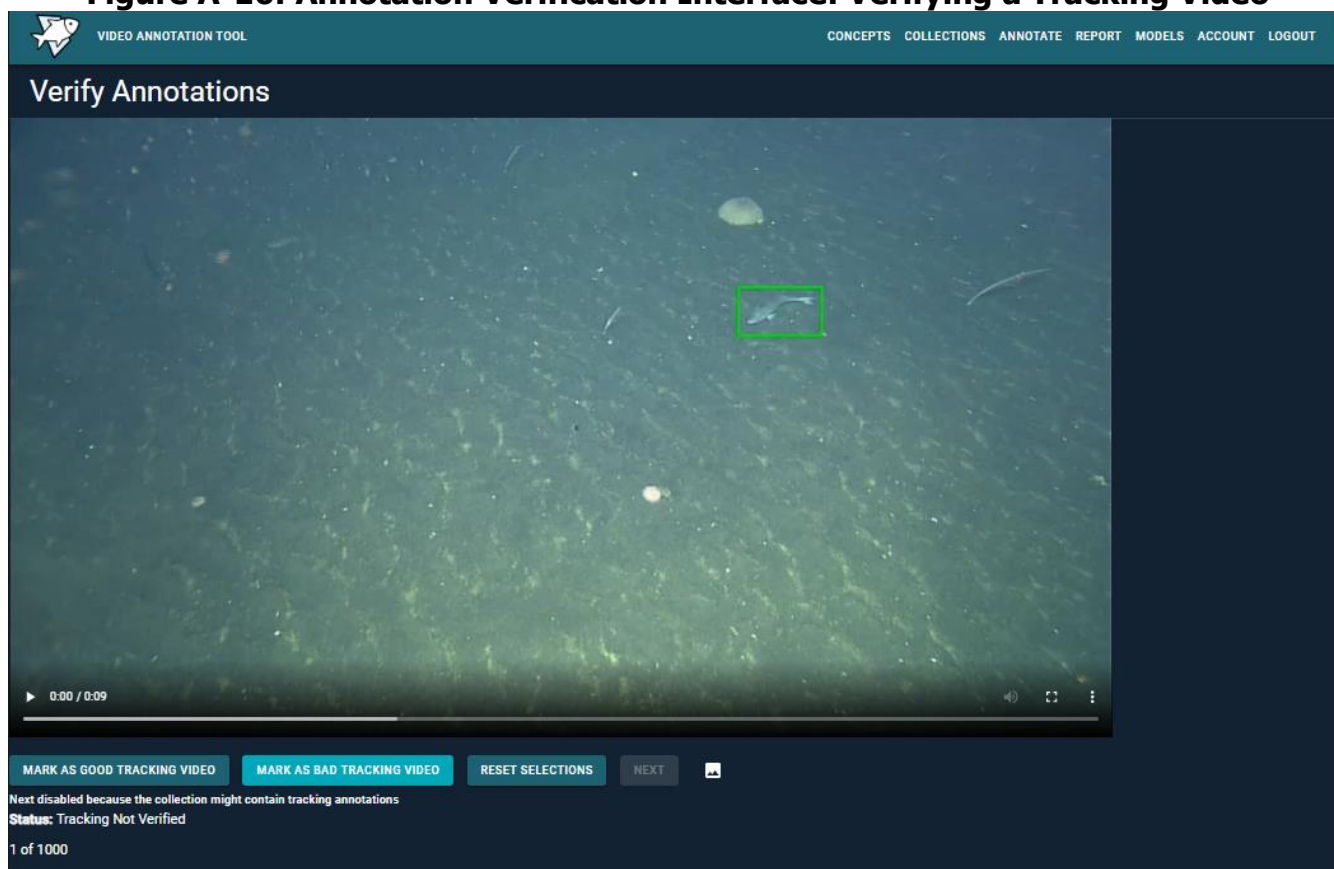


Shows how to verify an annotation in the Verification Interface.

Source: Cal Poly San Luis Obispo and MBARI

Red is used to display the annotation that one is hovering over with the mouse. This includes the option to delete the annotation. Green is used for user annotations that are already verified and are part of the collection. Blue is used for computer annotations that are outside the set of concepts for the annotation collection. Finally, orange is used for the bounding boxes of the annotations in the current frame that need to be verified. The **RESELECT SELECTIONS** button allows the user to reset the annotation collection that is currently being verified and gives them the option to select a new annotation collection to verify. The **DETAILS** button displays more information about the annotation, including the annotator, video, time in video, concept, and whether the annotator was unsure. It also allows the user to get detailed information about the video where the annotation was made, including GPS coordinates and depth of the start and end point, date and time when the video was recorded, and summary statistics about the video that includes the prevalence of different species in the video. The **DELETE** button deletes the current annotation. The **IGNORE** button moves to the next annotation and ignores the current one, while the **VERIFY** button marks the current annotation as verified before moving to the next one. There is also a **BACK** button that allows the user to go to previous annotation. The video image allows us to switch between an annotation and the tracking video of the annotation – see Figure A-16.

Figure A-16: Annotation Verification Interface: Verifying a Tracking Video



Shows how to watch and verify a tracking video in the Verification Interface.

Source: Cal Poly San Luis Obispo and MBARI

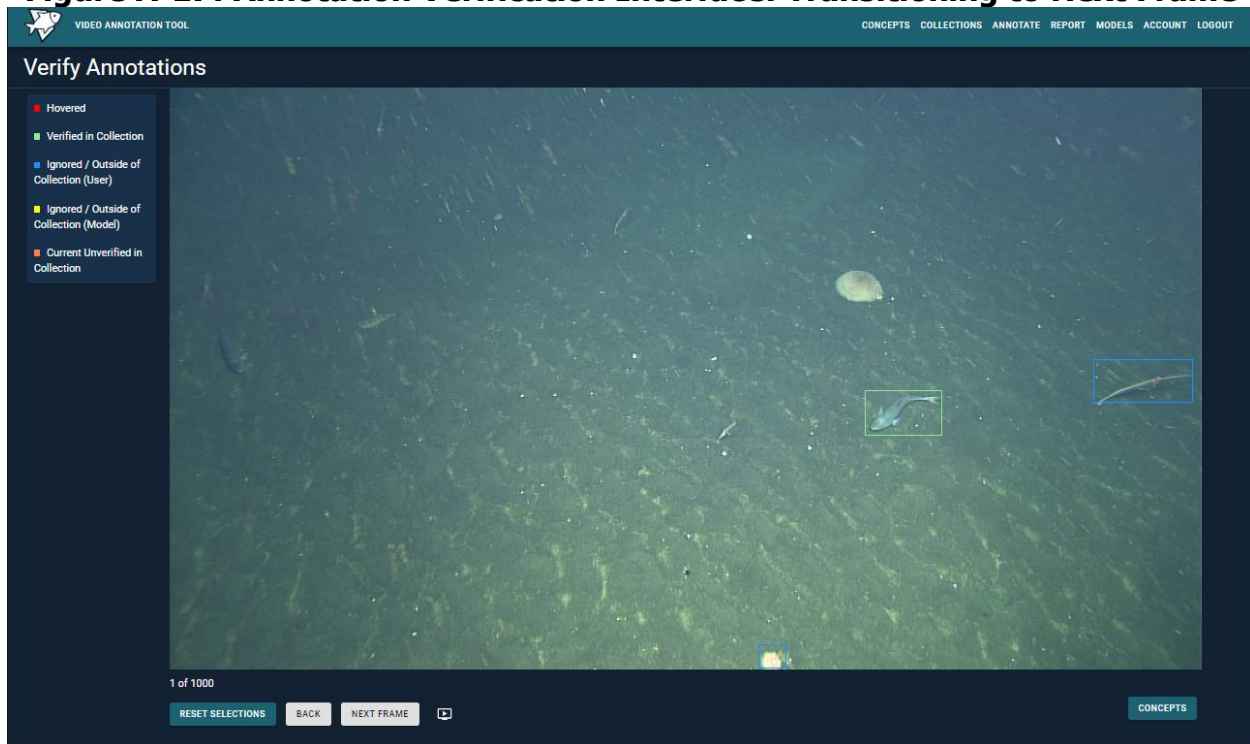
The buttons at the bottom allows us to mark the tracking video as good or bad, while the mountain image allows us to jump back to the frame verification menu. Note that the ability to see the tracking video of an annotation is very important because a single frame may not

suffice to determine if an annotation is correct. The `RESET SELECTION` button is included here as well, and it allows the user to select a different collection of annotations to verify.

If one wants to change the classification of a concept, they can select the bounding box of the concept, adjust it as needed, and then press the `CONCEPTS` button to open the concepts drawer and select the correct concept. Note that the current name and image of the concept in the selected bounded box appears below the video.

When one clicks the `VERIFY` button, a new screen for the same frame appears. It allows the user to enter additional annotations for the current frame – see Figure A-17. Note that the system relies on the fact that all species in the frames that are used for training a CNN are properly annotated.

Figure A-17: Annotation Verification Interface: Transitioning to Next Frame



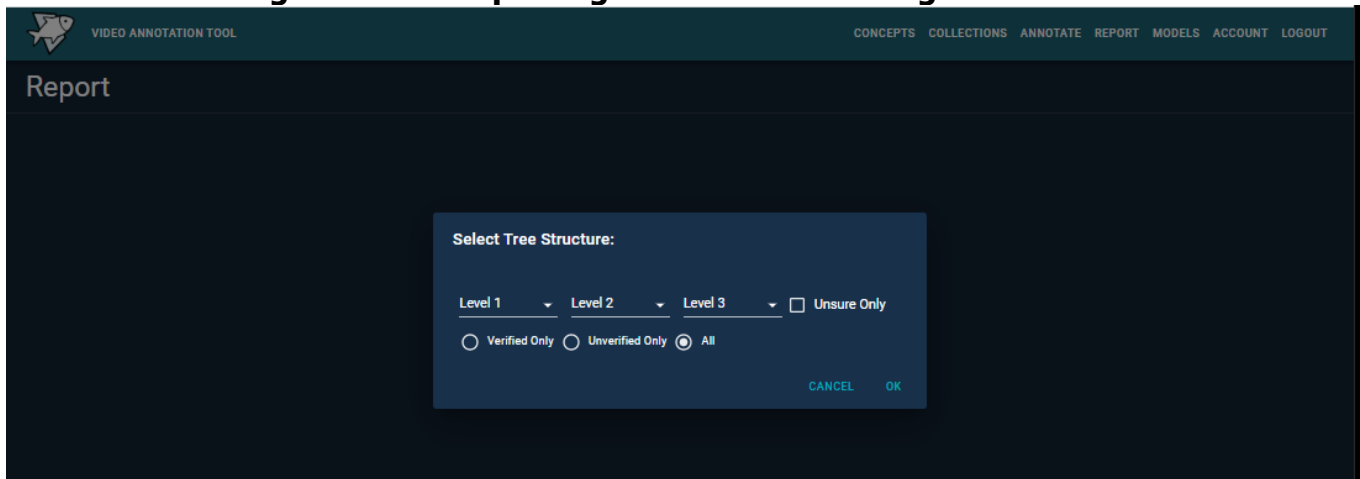
Shows an intermediate window before the user transitions to the next frame.

Source: Cal Poly San Luis Obispo and MBARI

REPORT Tab

The report tab displays non-tracking annotations. First one needs to choose the sorting order, where the annotations can be ordered by Video, Concept, or User. The menu also allows the user to select Unsure Only annotations and to select whether the displayed annotations should be verified or not – see Figure A-18.

Figure A-18: Reporting Interface: Selecting Sort Order

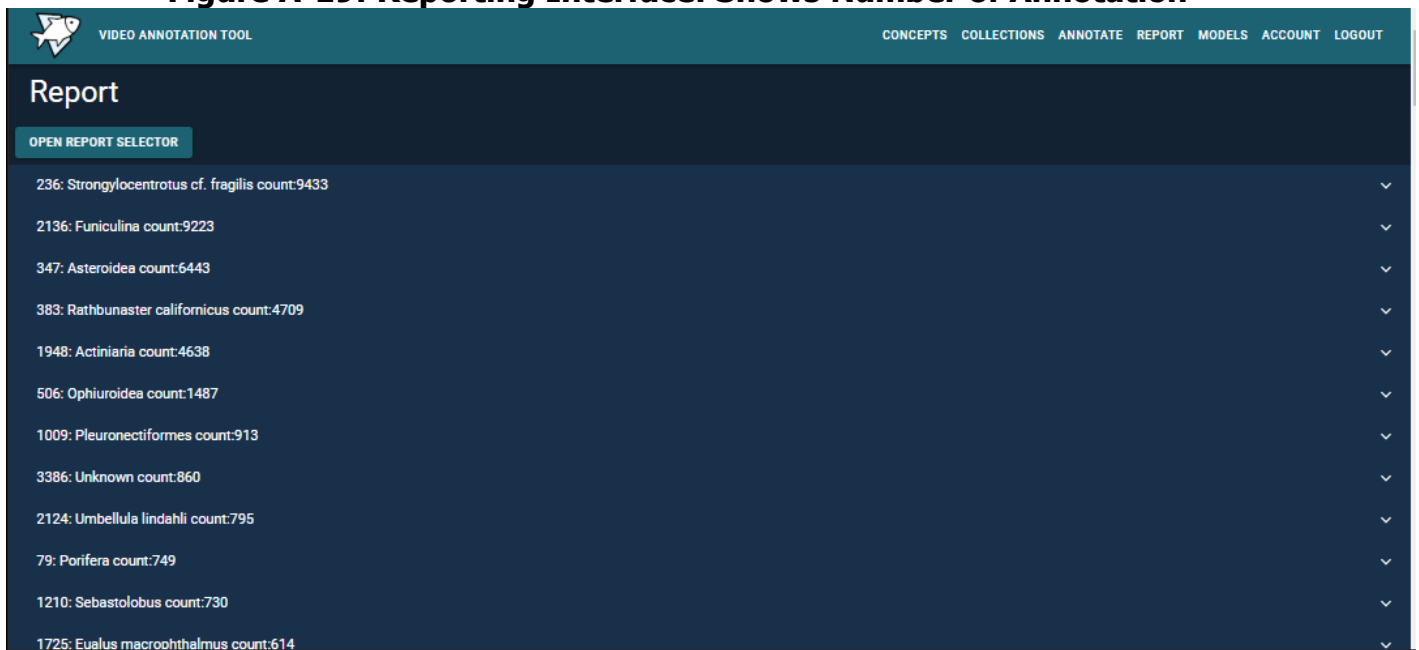


Shows how to select the sorting order in the reporting interface.

Source: Cal Poly San Luis Obispo and MBARI

Once the user makes their choice, all annotations are displayed in the selected order. Figure A-19 shows an example of sorting the annotations by concept first. On each line, the concept id, concept name, and total number of annotations is displayed. For example, one can see from the screenshot below that there are 9,433 annotations for the *Strongylocentrotus cf. fragilis*. One can use the up and down arrows on the right to navigate through annotations.

Figure A-19: Reporting Interface: Shows Number of Annotation

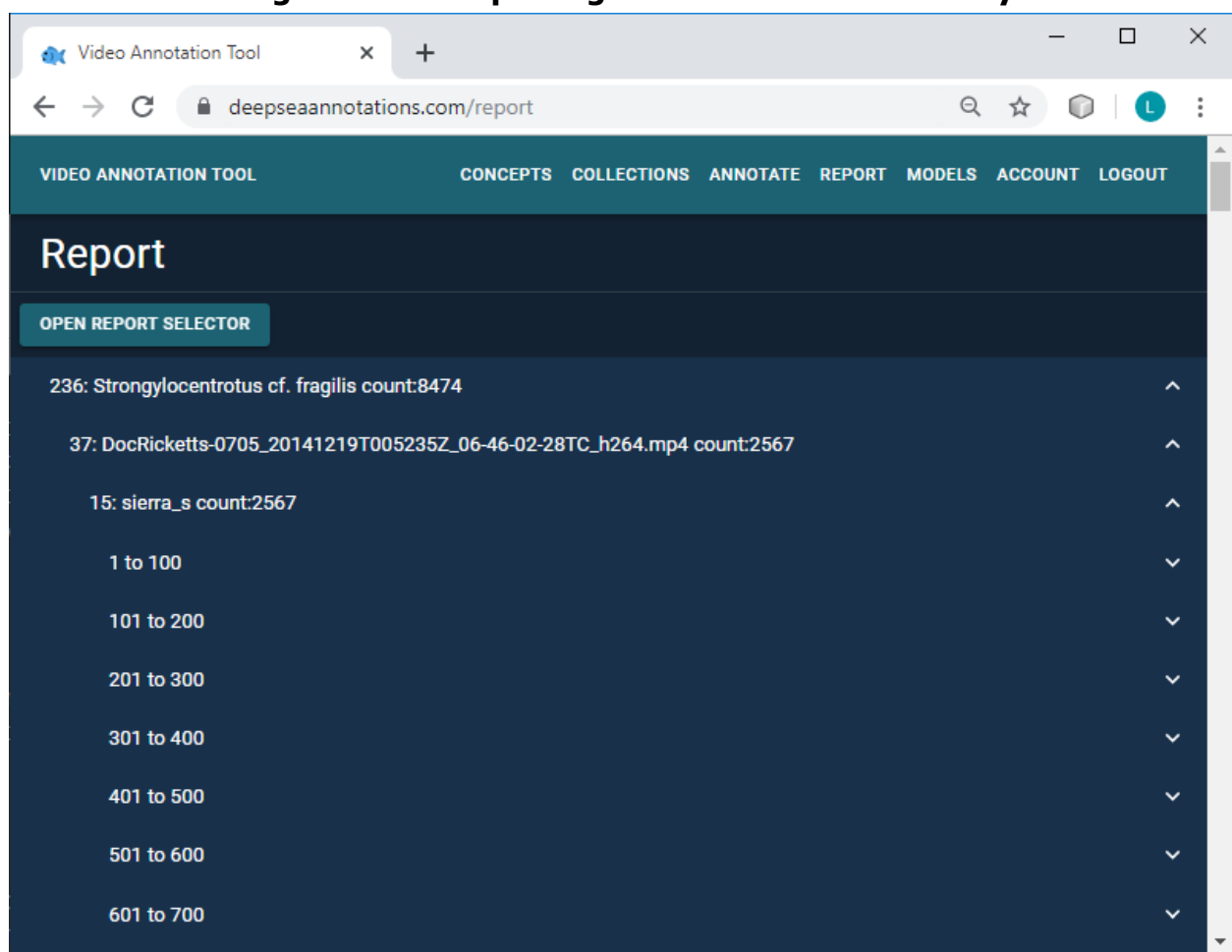


Shows a screenshot of the reporting interface that contains the number of annotations for each concept.

Source: Cal Poly San Luis Obispo and MBARI

The OPEN REPORT SELECTOR button on the top can be pressed at any time to reset the sorting order. The website displays no more than 100 annotations in each level of the display tree, where the concepts are split into buckets of 100s if there are more than 100 in any category – see Figure A-20.

Figure A-20: Reporting Interface: Tree Hierarchy

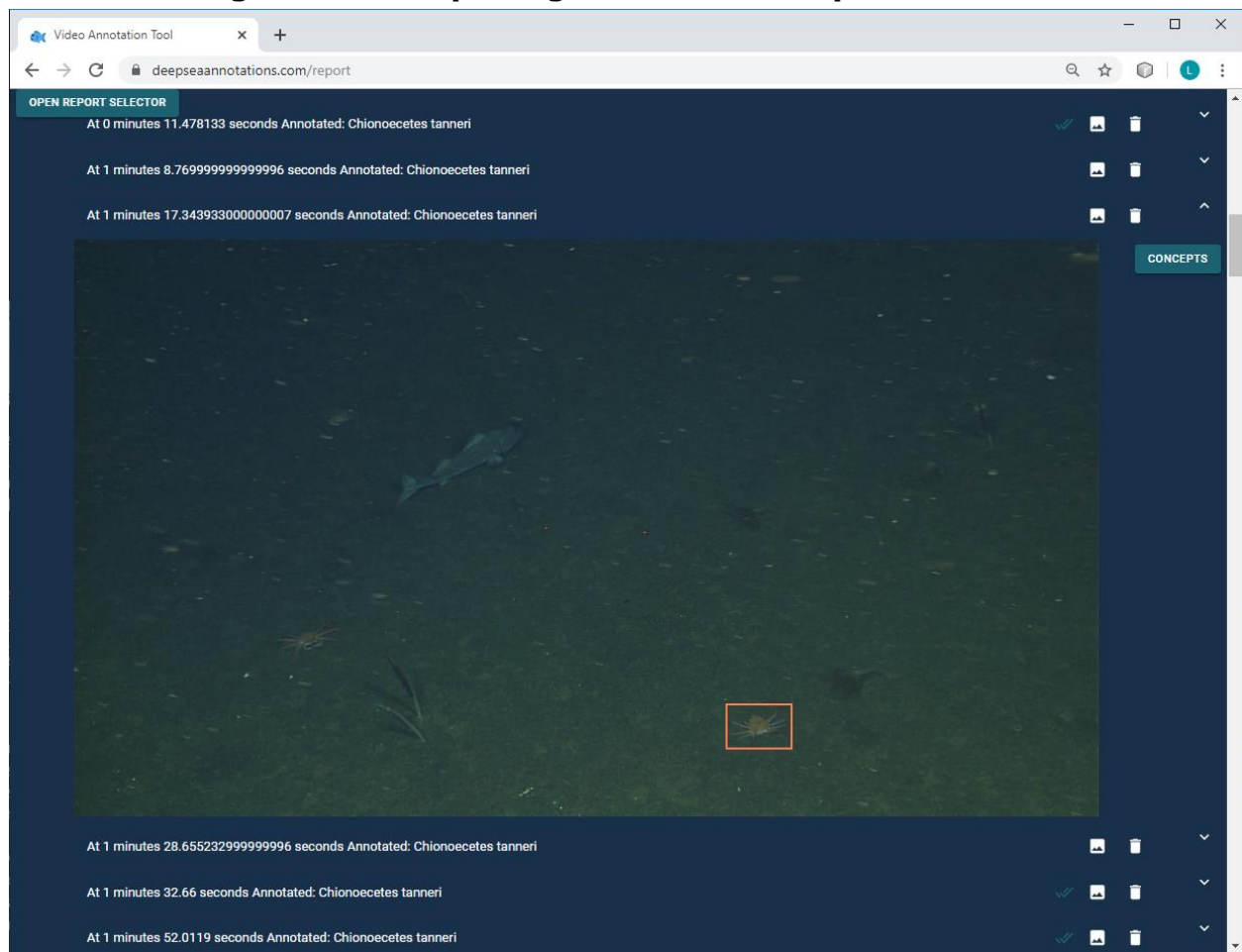


Shows the hierarchical display of the reporting interface.

Source: Cal Poly San Luis Obispo and MBARI

Once a specific annotation has been selected, one has the option of deleting it by pressing the delete button on the right, seeing the tracking video for it by clicking the video button, or changing the concept by clicking on the CONCEPT button, which opens the concept drawer – see Figure A-21.

Figure A-21: Reporting Interface: Example Annotation



Shows an example annotation in the reporting interface.

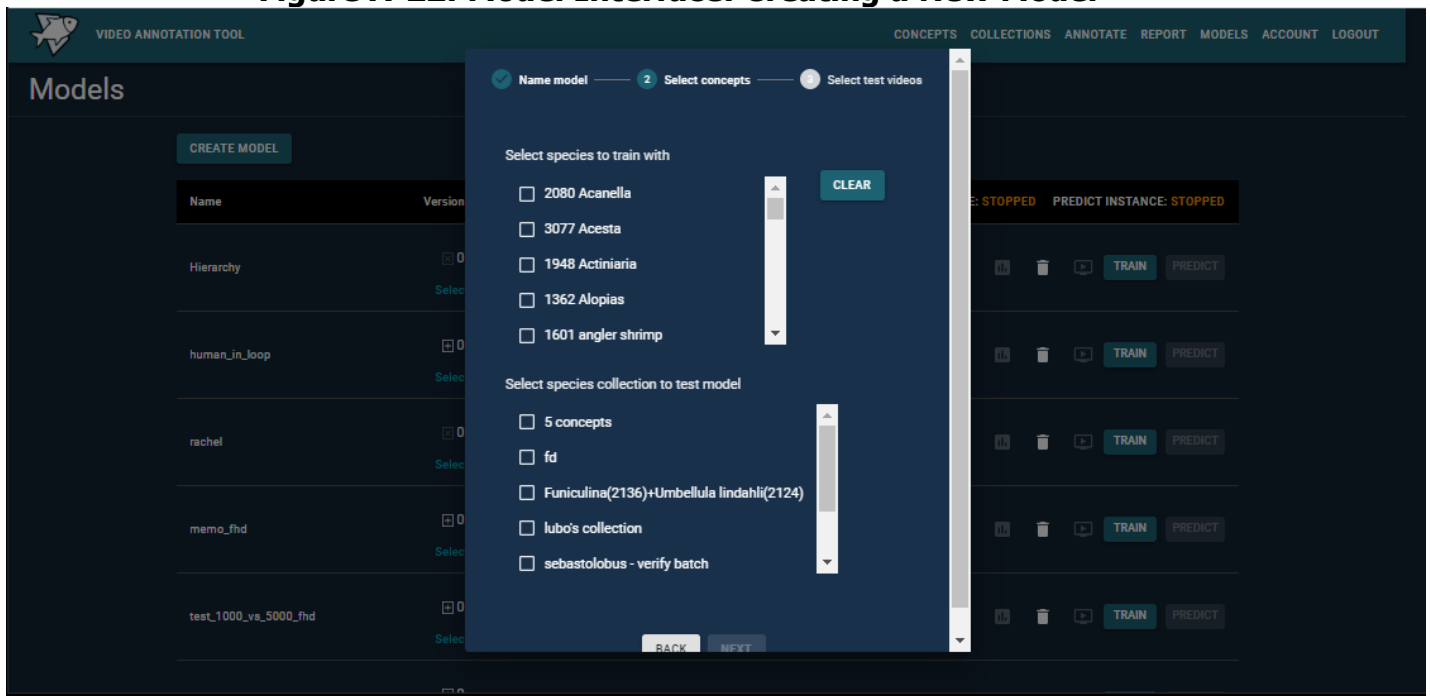
Source: Cal Poly San Luis Obispo and MBARI

Annotations that have been verified are displayed with double checkmarks on the right-hand side. The up and down arrows on the right-hand side can be used to expand and collapse the display tree.

MODELS Tab

To create a new model, one needs to click on the `CREATE MODEL` button at the top of the screen. This brings a three-step menu, where one can select the model name, the species to include in the model, and the test videos. The concept to be included can be selected from a list of concepts or a concept collection can be directly selected – see Figure A-22. Note that the system not only allows models to predict from a predefined set of concepts but also from concept collections. The idea is that when the CNN is not sufficiently confident which species is present, it can annotate it as a collection instead if the confidence is above the threshold.

Figure A-22: Model Interface: Creating a New Model



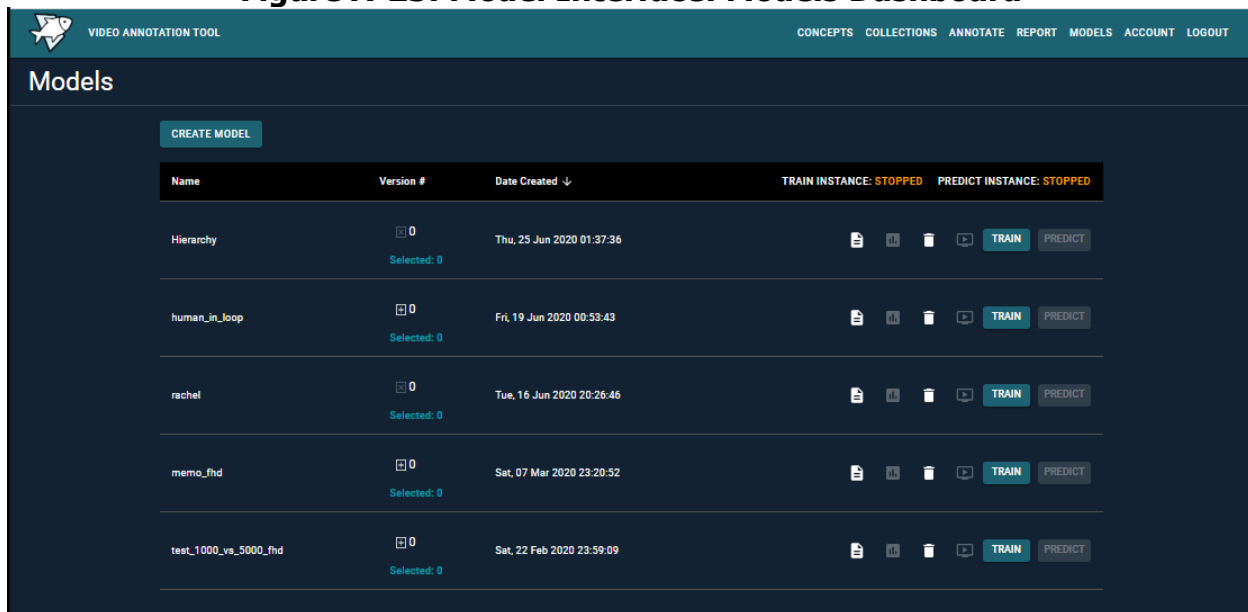
Shows an example of how to create a new model in the model interface.

Source: Cal Poly San Luis Obispo and MBARI

Lastly, one or more test (or verification) videos can be selected, where a video can be directly selected, or a video collection can be chosen. These videos will be used to verify the quality of a model and therefore the system will not train on annotations from them.

Once a model has been created, it will appear in the list of models and the only version number available will be 0. For each model, pressing the notes image will display information about the model; pressing the graphs image will display the Tensor Board for training the model; and pressing the delete image will delete the model. There is also a video button associated with every model that will show the computer-annotated verification videos for the model – see Figure A-23. Note that the top line of the table shows if there is currently a training or a prediction instance of the server running, where the software does not allow for multiple training jobs and multiple prediction jobs to run simultaneously.

Figure A-23: Model Interface: Models Dashboard



Shows the model dashboard in the model's interface.

Source: Cal Poly San Luis Obispo and MBARI

To create a new version of the model, one needs to press the `Train` button. A new popup window will appear, where one can specify the annotation collection to train on, the number of epochs, the number of images to use in the training, and whether to include unverified and/or tracking annotations – see Figure A-24. Note that the popup window includes a `TRAINING INFO` button, that tells the number and type of annotations the model version will train on – see Figure A-25.

Figure A-24: Model Interface: Training a New Model Version

The screenshot shows the 'Lubo's test' model interface. At the top, there are fields for 'Annotations' (1629), 'Epochs' (20), and '# of Images' (1200). Below these is a 'TRAINING INFO' section with two toggle switches: 'Include tracking annotations' (off) and 'Verified annotations only' (on). The main area is divided into three tabs: 'TRAINING STATUS' (selected), 'STANDARD OUTPUT', and 'STANDARD ERROR'. Under 'TRAINING STATUS', it shows 'Step 1/2' with the message 'Model has finished training.' and a 'RESET TRAINING' button. Below this, a progress bar shows 'Epoch: 4 / 10' and 'Batch: 104 / 104'. Further down, 'Step 2/2' shows 'Model has finished predicting.' and a progress bar for 'Video: 4 / 4' with 'frame 22500 of 22522'. At the bottom, the 'INSTANCE STATUS' is 'STOPPED'.

Shows how to create a new model version in the model's interface.

Source: Cal Poly San Luis Obispo and MBARI

Figure A-25: Model Interface: Training Info

Training Info					
Concept	User Annotations	Tracking Annotations	Verified User Annotations	Verified Tracking Annotations	All Annotations
Chionoecetes tanneri	317	0	220	0	317
Gnathostomata	1	0	0	0	1
TOTAL	318	0	220	0	318

Shows information about the annotations that will be used for training.

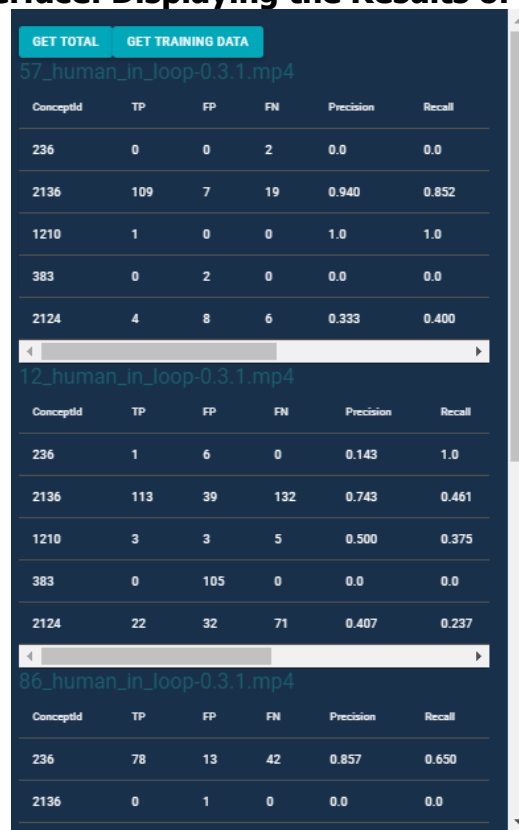
Source: Cal Poly San Luis Obispo and MBARI

Once the `START TRAINING` button is pressed, the window displays information about the training process.

Once the training is completed, a new version of the model is created. Note that the website creates a version tree. For example, clicking the `TRAIN` button for a model will create version 0.1 first and then version 0.2. Clicking the `TRAIN` button on version 0.1 will create first version 0.11.

Pressing the info image will give us information about the accuracy of the model version on the verification videos – see Figure A-26.

Figure A-26: Model Interface: Displaying the Results of Training a Model Version



ConceptId	TP	FP	FN	Precision	Recall
236	0	0	2	0.0	0.0
2136	109	7	19	0.940	0.852
1210	1	0	0	1.0	1.0
383	0	2	0	0.0	0.0
2124	4	8	6	0.333	0.400

ConceptId	TP	FP	FN	Precision	Recall
236	1	6	0	0.143	1.0
2136	113	39	132	0.743	0.461
1210	3	3	5	0.500	0.375
383	0	105	0	0.0	0.0
2124	22	32	71	0.407	0.237

ConceptId	TP	FP	FN	Precision	Recall
236	78	13	42	0.857	0.650
2136	0	1	0	0.0	0.0

Shows the accuracy of a model version on the verification videos in the model interface.

Source: Cal Poly San Luis Obispo and MBARI

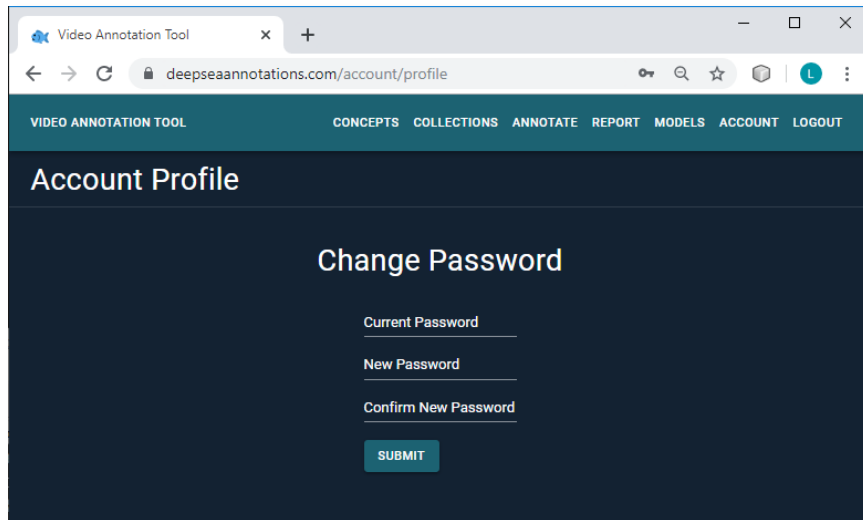
These numbers describe the quality of the video. One can get a better idea of how the model version performed by clicking the video button on the right. This will allow one to watch the automatically annotated videos with bounding boxes. In these videos, one can see concepts that are wrongly annotated and concepts that should have been annotated but were not. The trash icon can be used to delete a model version or a model.

Account Tab

The account tab has three menus: `Profile`, `Create Users`, and `Users`.

The `Profile` menu allows the user to change the current password – see Figure A-27.

Figure A-27: Account Interface: Changing a User's Password

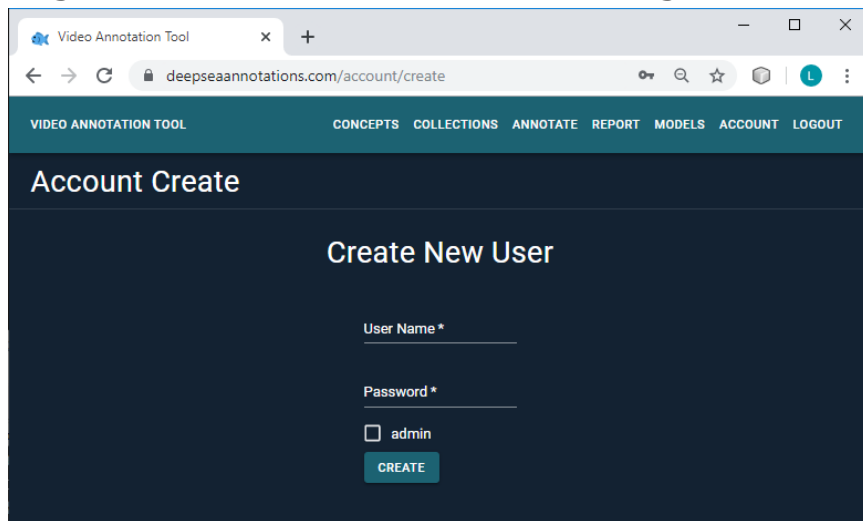
A screenshot of a web browser displaying the 'Account Profile' page of the 'Video Annotation Tool'. The page has a dark blue header with a navigation menu containing 'CONCEPTS', 'COLLECTIONS', 'ANNOTATE', 'REPORT', 'MODELS', 'ACCOUNT', and 'LOGOUT'. The main content area is titled 'Account Profile' and features a 'Change Password' section. This section includes three input fields labeled 'Current Password', 'New Password', and 'Confirm New Password', followed by a teal 'SUBMIT' button.

Shows how to change one's password in the account interface.

Source: Cal Poly San Luis Obispo and MBARI

The `Create User` menu allows us to create a new user, which can be an annotator or an admin. Only admin users have access to some of the functionality, such as training models – see Figure A-28.

Figure A-28: Account Interface: Creating a New User

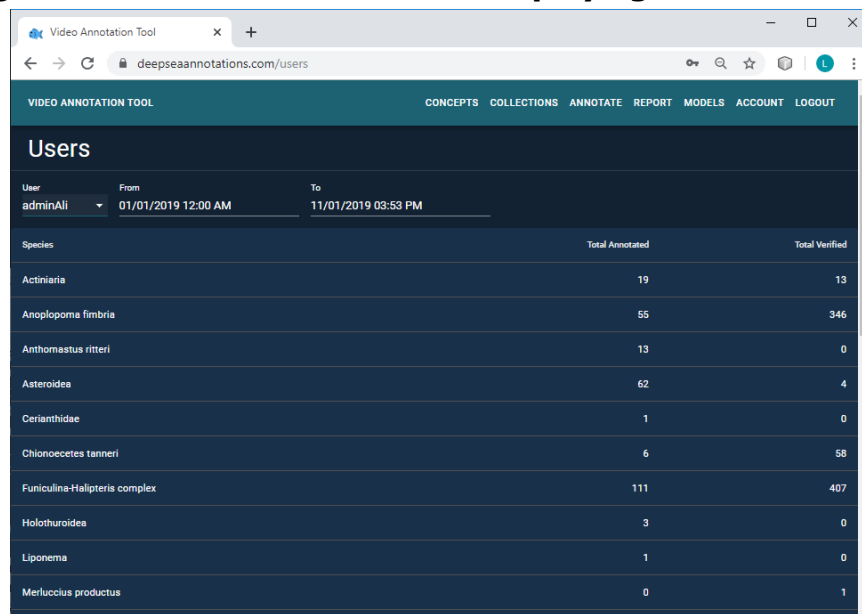
A screenshot of a web browser displaying the 'Account Create' page of the 'Video Annotation Tool'. The page has a dark blue header with a navigation menu containing 'CONCEPTS', 'COLLECTIONS', 'ANNOTATE', 'REPORT', 'MODELS', 'ACCOUNT', and 'LOGOUT'. The main content area is titled 'Account Create' and features a 'Create New User' section. This section includes two input fields labeled 'User Name *' and 'Password *', a checkbox labeled 'admin', and a teal 'CREATE' button.

Shows how to create a new user in the account interface.

Source: Cal Poly San Luis Obispo and MBARI

Lastly, the `Users` menu can be used to monitor the work of the annotators. Specifically, it can show the number of annotations for each user, concept, and time period – see Figure A-29.

Figure A-29: Account Interface: Displaying User Information



Species	Total Annotated	Total Verified
Actiniaria	19	13
Anoplopoma fimbria	55	346
Anthomastus ritteri	13	0
Astroidea	62	4
Cerianthidae	1	0
Chionoecetes tanneri	6	58
Funiculina-Halipteris complex	111	407
Holothuroidea	3	0
Liponema	1	0
Merluccius productus	0	1

Shows how to display the annotation and verification work done by the users.

Source: Cal Poly San Luis Obispo and MBARI

Website Use Cases

Next, possible users of the website and possible use cases are discussed, where the enumeration is not meant to be exhaustive. Note that the same user can potentially wear different hats and, in the very extreme, a single user can perform all tasks.

Administrator

The job of the administrator would be to create new accounts using the `ACCOUNT/ Create User` tab. They can also monitor the work of all the annotators through the `Account/Users` tab.

Annotator

The annotator will first create a bag of concepts from the `CONCEPTS` tab. Next, they go to the `COLLECTIONS/Videos` tab and create a collection of videos that they are interested in annotating. In creating this collection, they can check information about each video, including GPS coordinates, underwater depth, quality of video, has it been annotated or is it currently been annotated, and so on. Next, the annotator will go to the `ANNOTATE/Video` tab and start annotating concepts. The annotator can select videos from their video collection, where they need to verify that nobody else is annotating the video that they are annotating (a video that is already being annotated is shown in red).

The system will automatically remember the last video being annotated. This means that the next time the annotator logs in, they can just continue from where they left off. Note that an annotator can select a concept from the concept collection or add an additional concept as needed. If the annotator is not certain about the annotation, then they should still perform it, but mark it as `Unsure`.

As a user annotates a video, they can use the arrow keys on the keyboard to go forward/back in the video by one second, use the space key on the keyboard to stop the video, change the video play rate, or make the video full screen. Additionally, they can press the `TOGGLE CONTROLS` button to add/remove the control bar from the video.

Annotating videos is a task that requires particular skills and knowledge of species. An annotator can use the `REPORT` tab as a learning tool. In particular, it will show them all annotations of a particular concept, including tracking videos that show the type of movements that the species makes.

Annotator Supervisor

An annotator supervisor has several roles. First, they can watch the different videos and put them in different video collections. They can also mark a video as good quality or poor quality. Next, they need to identify the videos that need to be annotated and assign them to different annotators. After the videos have been annotated, the annotator supervisor should verify the annotations through the `ANNOTATE/Verify` tab. They can also use the `ACCOUNT/Users` tab to see the number of annotations done by each annotator in each time period.

Computer-Aided Annotator

The job of this type of user is to use the website to automatically annotate new unwatched videos. Experimental results have shown that in order to have a concept automatically annotated, there need be at least 1,000 human annotations of it. Having fewer annotations usually results in low count accuracy prediction.

In performing this task, the first job is to create a concept collection of the concepts that need to be automatically annotated. Next, one needs to use the `REPORT` menu to find the videos that contain these concepts. A good strategy is to select two or three videos for verification and the rest of the videos for training. Initially, the model should be trained on all the training videos and only the first verification video is to be used for verification. Experimental results have shown that verifying the annotations and using tracking data results in higher accuracy. Next, the computer-generated annotations on the verification video should be verified by a human and corrections should be made. After that, the model should be trained again, but this time the first verification video can be used as training video, where the annotations that were corrected are given higher weights. This process can be repeated one more time if there are three verification videos. In this way, the algorithm can learn from its mistakes.

Once a model is trained, the verification video is annotated automatically. A human can examine this video and see which occurrences of concepts are correctly annotated and which were not annotated or misclassified. Adding negative examples of similar concepts can be beneficial. For example, if the algorithm misclassifies sea stars as crabs, then adding annotations of crabs to the model will decrease the number of false positives.

In general, adding more annotations (of the concepts that one wants to classify or similar concepts), verifying more annotations, and increasing the number of epochs are good techniques to increase the accuracy of a prediction model.

APPENDIX B:

CNN Model Results

Experimental Procedure

Researchers conducted multiples experiments on the annotated data. They used 1280x720 pixel images from training the neural network (NN) model. They used an Amazon Web Services (AWS) g3.16xlarge, which has 4 NVIDIA Tesla M60 GPUs (32 Gigabytes GPU memory) to run the experiments.

When running an experiment, researchers first trained a CNN on selected species. Then they verified the quality of the model on verification videos, which contain human annotation that they did not train on. They ran multiple experiments with different species and hyperparameters. For each species, the following metrics describe the accuracy of the predictive results.

- TP = The number of true positives. This is the number of object occurrences that were classified by the AI algorithm the same as the human annotator.
- FP = The number of false positives. This is the number of object occurrences that were classified by the AI algorithm as one concept, but were not classified by the human or they were classified by the human annotator as a different concept.
- FN = The number of false negatives. This is the number of object occurrences that were not classified by the AI algorithm as a concept, but were classified by the human annotator as that concept.

$$P = \frac{TP}{TP+FP} . \text{ This is the precision.}$$

$$R = \frac{TP}{TP+FN} . \text{ This is the recall.}$$

$$F1 = \frac{2 \cdot P \cdot R}{P+R} . \text{ This is the F1 score.}$$

- MC = model count. This is the number of occurrences for a specific concept that were identified by the CNN model in a set of videos.
- UC = user count. This is the number of occurrences for a specific concept that were identified by a human annotator in a set of videos.
- CA = count accuracy = $1 - \frac{|MC-UC|}{\max(MC,UC)}$

The precision metric tells us how good are the annotations made by the algorithm. In other words, it checks in what percent of the time an annotation that is made by the algorithm is correct. Conversely, the recall metric tells us what percent of the human-made annotations were discovered by the algorithm.

It is easy to create an algorithm that has high precision, for example by only annotating few objects when the algorithm is 100% sure that an annotation is correct. Similarly, it is easy to create an algorithm with high recall by annotating everything the algorithm finds, even if it is doubtful that the annotation is correct. The difficult part is to create an algorithm that has

both high precision and recall. The F1 score is a combination of the two metrics. Achieving high F1 score means that the algorithm is generally good at annotating objects.

Note that all occurrences in a video of a concept are summed, assuming that the same organism will not appear multiple times in the same video (that is, no double counting). The reason is that an ROV moves at constant speed and projects a bright light that will make it highly unlikely for a creature to follow it. The fact that an ROV moves at a straight line also makes it unlikely that the same individual will occur multiples times in the same video.

Note that the videos that were used for training and predicting models are from MBARI. Some of the videos are good quality, while other videos were not shot at the ocean floor and therefore many targets are far away and hard to identify. For this reason, researchers marked some of the videos as good videos.

Experimental Results

The researchers conducted several experiments. Note that some of the experiments include the concepts from the Species Selection Memo, while in other experiments, different concepts were identified.

Star Fish and Urchin

Experiments were run on videos with (1) *Rathbunaster californicus* (starfish) and (2) *Strongylocentrotus fragilis* (sea urchin) were considered because of their abundance in the videos and ecological importance. The results are shown in Tables B-1 and B-2.

Table B-1: Top Concepts Trained with Human Data

Concept	TP	FP	FN	P	R	F1	MC	UC	CA
1	134	11	9	0.924	0.937	0.931	145	145	1.00
2	77	12	35	0.865	0.688	0.766	89	140	0.636

Results of training the CNN model on 5,000 annotations per concept from human annotators exclusively (that is, just the frames that a person annotated).

Source: Cal Poly and MBARI

Table B-2: Top Concepts Trained with Human and Generated Data

Concept	TP	FP	FN	P	R	F1	MC	UC	CA
1	126	14	13	0.900	0.906	0.903	140	145	0.966
2	109	30	21	0.784	0.838	0.810	139	140	0.993

Results of running the same experiment, but also including tracking data (multiple frames for each occurrence).

Source: Cal Poly and MBARI

Other Species Concepts





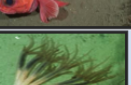
These experiments were run on five additional species, including:

1. *Strongylocentrotus fragilis* (deep water sea urchin),
2. *Funiculina spp.* (sea pen),

3. *Sebastolobus spp.* (thornyhead rockfish),
4. *Rathbunaster californicus* (deep water sun star), and
5. *Umbellula lindahli* (sea pen).

5,000 annotations per concept were used. The results are shown in Table B-3.

Table B-3: 5,000 Annotations Per Concept, Good Data

	Name	F1 score	Predicted	Actual	Count Accuracy
	Sea urchin	0.667	79	53	0.639
	Sea pan	0.610	46	54	0.852
	Sunflower star	0.772	49	72	0.681
	Rockfish	0.729	47	53	0.887
	Sea pen flower	0.762	11	10	0.901

Results of running experiment with 5,000 annotations per concept. For concepts that had less than 5,000 human annotations, researchers included annotations from tracking. Only data from videos with good image quality was used.

Source: Cal Poly and MBARI

Analyses of Experimental Results

As expected, model predictive accuracy increases with larger numbers of training annotations.

Note that the annotations are not complete (i.e., not every organism in the videos could be annotated). In some cases, videos have frames with hundreds of concepts, so it is understandable why they are not all annotated by the human annotator. In other cases, the annotators just missed annotating some frames, which cannot be corrected through the verification menu. Missing annotations results in lower precision, but it doesn't affect recall. This is the reason why the precision numbers were low in some experiments. Researchers have also found that the CNN struggles with correctly annotating small objects, such as sea urchins. This is a common problem in computer vision.

In summary, the experimental results have shown that the website can produce annotations with reasonable accuracy, at least for the concepts tested. This serves as a proof of concept that an AI approach is reasonable. However, more work needs to be done to expand the database of annotations for more concepts, improve the accuracy of the models, and make the website commercially viable.