Check for updates

FEATURE

Keeping Track of Hawaii's Bottomfish Populations With the Help of Citizen Scientists

Hana Ra | University of Hawaii, Hawaii Coral Reef Initiative, 2424 Maile Way, #718, Honolulu, HI 96822

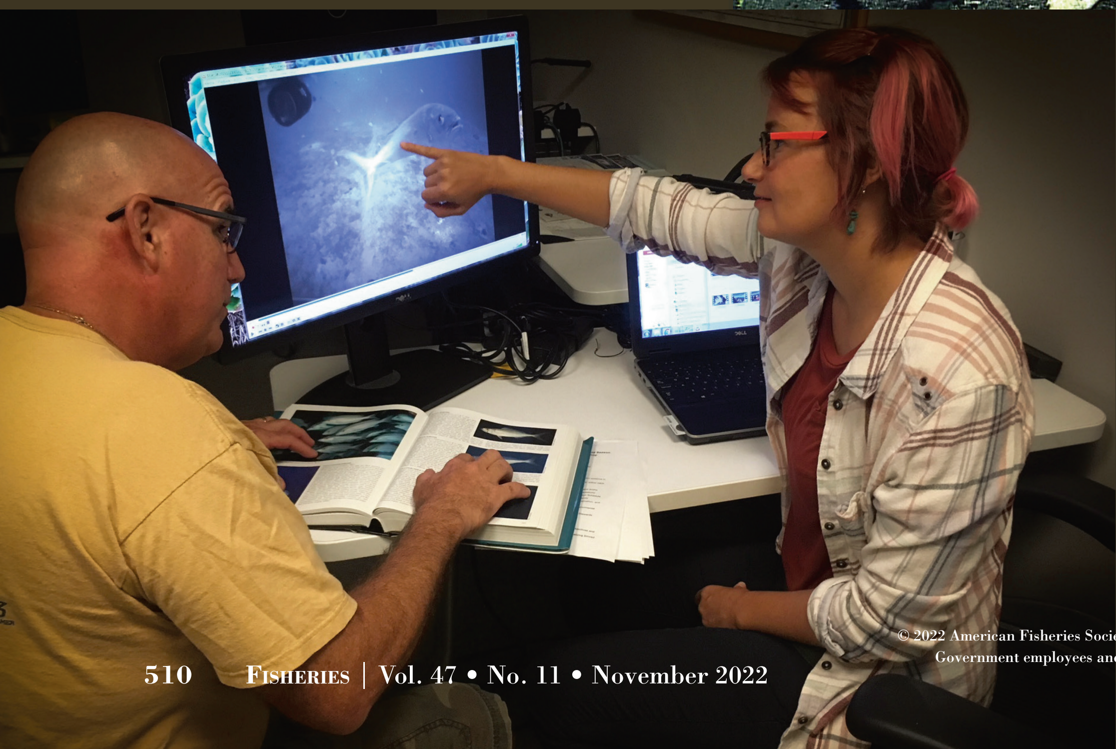
Benjamin L. Richards | National Oceanic and Atmospheric Administration Pacific Islands Fisheries Science Center, Honolulu, HI.
E-mail: benjamin.richards@noaa.gov

Audrey Rollo | Cooperative Institute for Marine and Atmospheric Research, Honolulu, HI

Dianna Miller-Greene | Cooperative Institute for Marine and Atmospheric Research, Honolulu, HI

Jeremy Taylor | Cooperative Institute for Marine and Atmospheric Research, Honolulu, HI

Onaga Etelis coruscans photographed as part of the OceanEYEs project. Photo credit: NOAA Office of Ocean Exploration and Research.



Christopher Demarke (left) and Dianna Miller-Greene (right) from the NOAA Pacific Islands Fisheries Science Center review images captured through the OceanEYEs progam. Photo credit: National Marine Fisheries Service.

DOI: 10.1002/fsh.10812

The “Deep 7” bottomfish complex, which consists of six snapper and one grouper species, is a complex that carries high economic and cultural importance to the islands of Hawaii. These bottomfish have been monitored through the National Oceanic and Atmospheric Administration Pacific Islands Fisheries Science Center’s Deep 7 fishery-independent surveys since 2016. These surveys use underwater stereo camera systems that produce hundreds of thousands of images that must be annotated by human analysts in order to generate species-specific, size-structured abundance estimates. We developed a citizen science project, called “OceanEYEs,” as a means to effectively process this imagery. A beta test was conducted to determine the accuracy of citizen science annotations in comparison to expert annotators. Our results suggest that aggregated citizen scientist data can achieve accuracy levels approaching that of expert annotators, which has the potential to improve image annotation efficiency and produce large volumes of high-quality training data to improve machine learning algorithms.

BACKGROUND

Commercial and recreational fishing have been important components of the Hawaiian culture and economy for hundreds of years (Haight et al. 1993). A complex of six deepwater snappers and one grouper species, commonly known as the Hawaii “Deep 7” (Western Pacific Regional Fishery Management Council 2010; Figure 1), are of central importance and have been under a formal federal management plan since 2005, when it was determined that the stock was experiencing overfishing (Moffitt et al. 2006).

The National Oceanic and Atmospheric Administration (NOAA), in partnership with the state of Hawaii, is responsible for actively managing the Deep 7 to ensure fishery sustainability. Formal stock assessments are routinely conducted by NOAA’s Pacific Islands Fisheries Science Center (PIFSC; Langseth et al. 2018) to determine the status of the Deep 7 relative to certain management-determined reference points. These assessments require reliable estimates of life history demographics, fishery catch, and population abundance. In 2016, to improve the abundance metrics used in the stock assessment, the PIFSC initiated a Bottomfish Fishery-Independent Survey in Hawaii (BFISH; Richards et al. 2016).

The BFISH is conducted annually from July to November. The survey domain (Figure 2) spans 600 km from Ni’ihau to the Big Island of Hawaii, comprising approximately 6,500 km^2 and covering the 75–400-m depth range around all eight main Hawaiian islands (Richards et al. 2016). The survey domain is gridded at 500 m, with each grid assigned to habitat-based design strata. Each year, a subset of survey grids is selected for sampling using a stratified random experimental design and one of two calibrated (Richards et al. 2016) survey gears—research fishing or the Modular Optical Underwater Survey System (Amin et al. 2017)—is assigned to each survey grid.

Within each assigned grid, two replicate 15-min camera samples are collected. With a typical survey comprising up to 100 camera grids, nearly 50 h of stereo video, comprising 2.2 million image pairs may be generated. It typically takes a team of three human annotators 2–3 months to process this volume of data using the MaxN methodology outlined by Cappo et al. (2006). With a limited analyst pool, this methodology is not scalable to future projected data volumes, nor does it allow for the image-by-image bounding box-style annotation required to train novel machine learning algorithms properly (Richards et al. 2019). Nor is it conducive to investigating alternative enumeration metrics, such as MeanCount, which have been shown to be desirable in some domains (Schobernd et al. 2013; Campbell et al. 2015).

With the widespread use and success of citizen science platforms (Kosmala et al. 2016) for processing imagery across a wide range of domains (Raddick et al. 2010; Cox et al. 2015), PIFSC has developed a citizen science tool, called “OceanEYEs,” using the Zooniverse platform, a citizen science web resource created by the Citizen Science Alliance which hosts some of the largest and most successful citizen science projects (Simpson et al. 2014).

PROJECT DESCRIPTION

The OceanEYEs project was developed to allow volunteer citizen scientists to help annotate underwater images from the BFISH survey, enabling researchers to investigate new enumeration methods, and to produce high-quality training data for machine learning algorithms. Volunteers are guided through the classification workflow, in which a series of tasks are completed that provide data on fish presence/absence, fish count, fish position, and species identification. A variety of tools are provided to help users before and during the workflow,

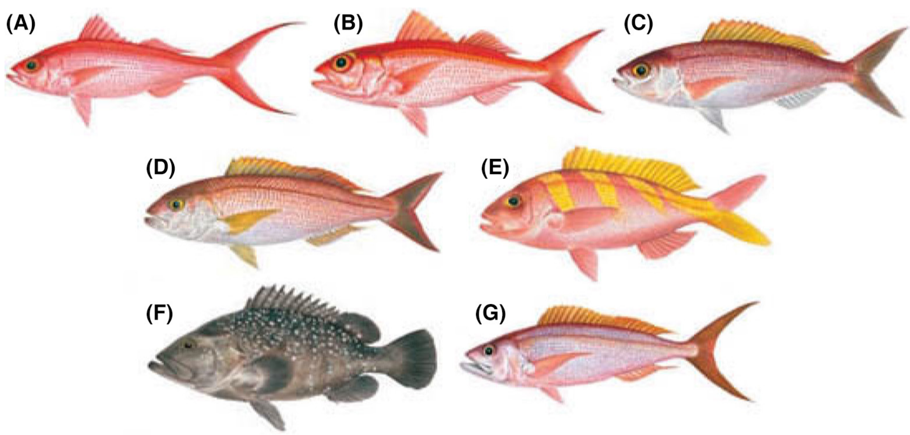


Figure 1. The main Hawaiian Islands “Deep-7” bottomfish complex: (**A**) Onaga *Etelis coruscans*, (**B**) Ehu *Etelis carbunculus*, (**C**) Kalekale *Pristipomoides sieboldii,* (**D**) Opakapaka *P. filamentosus*, (**E**) Gindai *P. zonatus*, (**F**) Hapu’upu’u *Hyporthodus quernus*, and (**G**) Lehi *Aphareus rutilans*. Artwork credit: Les Hata, Hawaii Department of Land and Natural Resources.

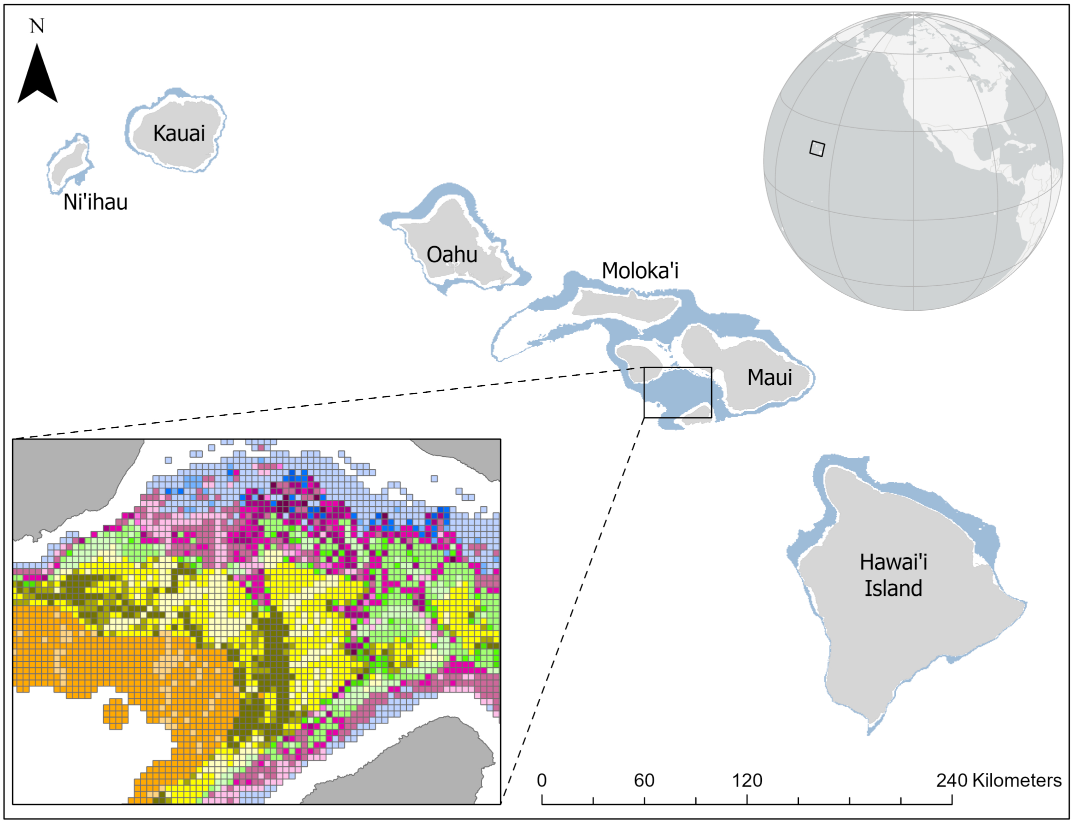


Figure 2. The spatial frame of the Bottomfish Fishery-Independent Survey in Hawaii (BFISH) survey extending from Kauai in the northwest to the island of Hawaii in the southeast. Inset shows a section of the survey frame in the Maui–Nui region showing the 500- × 500-m grid cells classified by habitat–depth strata.

including a tutorial on how to complete the workflow, a field guide that provides additional information on each of the Deep 7 bottomfish species, and a frequently asked questions section that can help users with any additional questions they may have regarding the workflow and the project overall.

To measure the usability and effectiveness of the OceanEYEs project in collecting high-quality annotation data, a test to validate the project was conducted in the summer of 2019, where users were asked to complete the workflow that was developed. This beta test consisted of 100 underwater images, with 70 images containing fish and 30 images containing no fish. A total of 208 users participated in this beta test, and each beta test image was annotated by 15 independent volunteers, a cutoff that was determined randomly, before it was "retired" from the workflow. Once the beta test was complete, data was extracted and compiled into consensus annotations using a graphical user interface provided by Zooniverse and python scripts. These tools required a few parameters to conduct the extraction and reduction process, including "eps," a threshold distance between two points, and "counterfrac_limit," a limit based on the fraction of classifications made for a particular image. Thresholds for these parameters were tested by analyzing nine different parameter combinations to assess which combination yielded the highest overall accuracy. Accuracy of the data was determined by comparing users' annotations to ground truth annotations that were provided by a PIFSC expert video analyst.

The four main data criteria of interest—fish presence/absence data, count data, positional data, and species identification data—were analyzed sequentially from most to least important. The most important criterion was fish presence/absence data, which was analyzed by organizing inaccuracies into two categories: "false positive," which is when the consensus annotations resulted in no fish present for an image when there was actually fish present, and "false negative," which is when consensus annotations resulted in fish being present in an image, when there was actually no fish present. Results for fish presence/absence data showed false positives only occurred in less than 8% of all analyzed photos throughout the nine parameter combinations and false negatives only occurred in less than 1% of the photos analyzed throughout all nine parameter combinations.

Count data, the second main criteria of interest, was analyzed by assessing the total amount of images that were overcounted and the total amount of images that were undercounted for each of the nine parameter combinations (Table 1). Results for count data showed the highest accuracy being associated with an eps parameter of 50 and a counterfrac_limit parameter of 0.2, which had an overall overcounting of one image and overall undercounting of 15 images, equating to an overall accuracy of 84%.

The third criterion of interest, positional data, was analyzed by visually evaluating the accuracy of the point placement throughout output images that were generated by the python scripts (Figure 3). The accuracy of the positional data was organized by dividing points into two main categories: "correct"/"on position" points, which are points that are correctly placed on the head and tail of the fish, and "incorrect"/"off position" points, which are points that are not correctly marked on the head or tail of a fish. Incorrect/off

Table 1. Results of beta test fish count data assessed throughout nine parameter combinations, including data on the number of images that were overcounted, the number of images that were undercounted, and the total proportion of images that were counted correctly.

	eps 50; cf 0.2	eps 100; cf 0.2	eps 200; cf 0.2	eps 50; cf 0.3	eps 100; cf 0.3	eps 200; cf 0.3	eps 50; cf 0.4	eps 100; cf 0.4	eps 200; cf 0.4
Overcounting	1	3	4	1	2	1	0	1	0
Undercounting	15	22	36	21	26	38	28	31	41
Proportion correct	0.84	0.75	0.60	0.78	0.72	0.61	0.72	0.68	0.59

position points were divided into 2 subcategories: "opposite" points, in which the head point is on the tail of a fish or the tail point is at the head of a fish, and "rogue" points, which are points that are not correctly placed on either the head or the tail of the fish (Figure 4). Using this classification scheme, results showed that with the optimal parameter combination of eps = 50 and counterfrac_limit = 0.2, the overall accuracy of positional data was 88.8%.

Species identification data was analyzed by comparing the consensus annotation species identification generated by user annotations to the ground truth species identification evaluated by the expert video analyst. This data was then summarized into a confusion matrix (Table 2) and then transformed as a ratio of correctly identified fish for a particular species over the total of the particular species present, excluding unmarked fish, to quantitatively evaluate the accuracy of the species identifications made from the beta test (Table 3). Results for this portion of the beta test analysis showed an overall accuracy of around 75.4%.

The overall results for the beta test show that when using the optimal parameter combination, fish presence/absence data resulted in a 97% accuracy, count data resulted in an 84% accuracy, an 89% accuracy for positional data, and a 75% accuracy for species identification data; these accuracies average to an 86% accuracy, which is comparable to expert annotator accuracy levels.

Since the project has become open to the public, OceanEYEs has amassed nearly 2 million annotations spanning over 160,000 images with the help of over 10,000 registered Zooniverse volunteers. Analyses of the annotation data gathered since the public launch have shown that the accuracy of the annotation data has decreased. This decline in accuracy is in part because the beta test intentionally selected images to ensure target fish were visible, resulting in fish that were more identifiable than typically seen due to their proximity to the camera. In contrast, the public launch batches consist of photos taken throughout the entire deployment, with fish occurring in various positions (e.g., in the distance or camouflaged in the substrate). Fish in these configurations could easily be missed or misidentified since key anatomical features are harder to see at a distance. To minimize the impact that fish at a distance can have on identification, a new feature is being added to the project, in which an additional second photo is included in the workflow to assist in spotting these occurrences. This second image will enable the users to detect movement, as well as move a fish into a more identifiable position (e.g., if the fish was obstructed by another object or if the fish was positioned straight towards the camera), which may help in increasing the accuracy of user annotations. Additionally, a new section will be added to the field guide addressing the differences between Opakapaka *Pristipomoides filamentosus* and Kalekale *P. sieboldii*, two Deep 7 bottomfish species that are difficult to differentiate. This new section will hopefully serve to help increase the accuracy of species identifications made by users, especially between these two species.

IMPROVING THE SCIENCE

Artificial intelligence, computer vision, and machine learning continue to expand into the marine science domain and are revolutionizing the way scientists collect and process data. The National Marine Fisheries Service and other agencies are increasingly turning to camera-based instrumentation to survey resource populations. While such sensors greatly increase the efficiency of field data collection, much of that efficiency gain is lost in the human-based image processing workflow. New machine learning methods based on artificial neural networks, commonly known as "deep learning" methods, are beginning to reduce the human annotation burden, but such methods require extensive libraries of human-annotated bounding box-style training data. With human analysts already fully tasked with image processing and data analysis, often using methods that do not produce frame-by-frame, bounding box-style annotations, production of this type of training data often remains a significant bottleneck. The Zooniverse platform

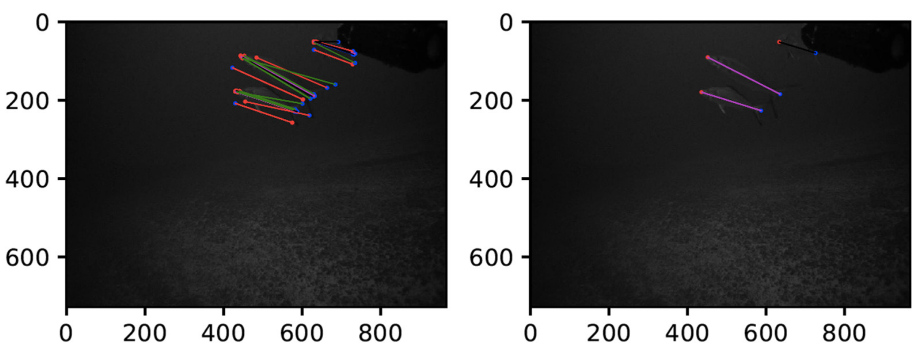


Figure 3. Example of an annotated image output generated from the python scripts. The left image contains annotations that all volunteers made for a particular image. The right image contains the associated consensus annotations created from the user annotations. Red dots in the image represent head points, blue dots represent tail points, and the different colors for each line represent the species identification (i.e., each of the Deep 7 species and "other" fish).

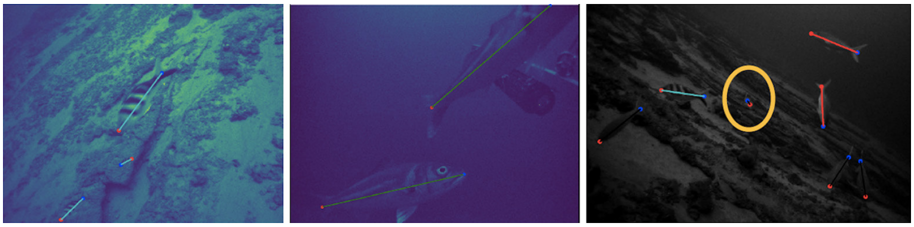


Figure 4. Examples of annotation categories created to assess the beta test positional data accuracy. The left image shows "correct"/"on position" points, with head points (red dots) and tail points (blue dots) For Peer Review Only correctly placed on each fish in the image. The middle image provides an example of "opposite" points, with the head points and tail points on the opposite ends of the fish. The right image shows an example of "rogue" points, in which head and tail points are not correctly placed on either the head or tail of the fish (points are circled in orange).

provides a convenient and easy-to-use interface for citizen scientists to create frame-level, bounding box-style annotations quickly and efficiently. These annotations can then be used to train machine learning models to continue the transition from human to machine-based image processing.

As each image is annotated by 15 independent citizen scientists before consensus annotations are generated, a confidence score based on the level of agreement among those independent annotators is assigned to each annotation. Adjudication and scoring of OceanEYEs annotations against professional annotations can be used to define a confidence threshold above which annotations are accepted with minimal verification and below which annotations are verified by professional annotators. In so doing, the OceanEYEs projects can reduce the overall burden on professional annotators and increase the overall efficiency of image processing.

Species are currently enumerated from BFISH video data using the MaxN method (Cappo et al. 2006). However, alternative enumeration metrics, such as MeanCount, have been shown to be desirable in some domains (Schobernd et al. 2013; Campbell et al. 2015). The frame-level annotations provided by OceanEYEs citizen scientists will allow researchers to investigate and compare alternative enumeration metrics to determine optimal methods for specific target taxa.

Finally, we hope that the education and outreach provided through the OceanEYEs platform will help users better understand the scientific process and data used to inform fisheries management. Current events have, yet again, shown the disconnect between the scientific community and the general populace. It is our hope that OceanEYEs can serve as one small step toward community engagement in the scientific process and can catalyze a renewed understanding and appreciation of the scientific process, its results, and implications. Effective management action is predicated largely on voluntary compliance, which is best achieved through understanding and acceptance.

LESSONS LEARNED/BEST PRACTICES

A variety of supplementary materials were provided to help users complete the OceanEYEs workflow, with particular emphasis on instructions regarding how to mark the fish and how to identify each of the Deep 7 target species versus other fish. When creating and modifying these materials, we learned that it is important to provide information that is educational and interesting to users, but is also not extraneous as to unnecessarily inundate volunteers. For example, one of the supplementary materials of focus was the field guide, which includes additional information regarding how to identify and mark each of the Deep 7 species and other fish found in the images. Although it would have been interesting and informative to include all of the potential fish that could be seen in these images in the field guide, having that much information would risk oversaturating the volunteers that are simply trying to find the fish they are interested in identifying. Therefore, since the target fish for this project are the Deep 7 bottomfish, with all other fish considered "other," we limited the field guide to include information on the Deep 7 species as well as any other fish that may be misidentified as one of the Deep 7 species. This way, volunteers only need to look through a small set of photos to find the specific species they may be looking for.

As was mentioned in the Project Description section or this article, a beta test was conducted to assess the accuracy and

Table 2. Confusion matrix of species identifications resulting from beta test. Column values represent "Truth" species identifications recognized by a video analyst expert, and row values represent "Predicted" species identifications resulting from the beta test consensus annotation output. The row labelled "Missing" represents fish that were unmarked in the consensus annotation output for each of the species of interest. Values where the species identifications in the row and column are identical represent the number of times consensus annotations correctly identified a particular species present.

	Ehu	Opakapaka	Kalekale	Gindai	Onaga	Lehi	Hapuupuu	Other
Ehu	11	NA	NA	NA	NA	1	NA	NA
Opakapaka	NA	20	10	NA	NA	1	NA	9
Kalekale	NA	NA	25	NA	NA	NA	NA	3
Gindai	NA	NA	NA	12	NA	NA	NA	NA
Onaga	NA	NA	NA	NA	3	NA	NA	NA
Lehi	NA	NA	NA	NA	NA	30	NA	3
Hapuupuu	NA	NA	NA	NA	NA	NA	8	NA
Other	NA	5	32	NA	2	9	2	41
Missing	12	NA	6	NA	NA	NA	2	12

Table 3. Results of beta test species identification data for each of the Deep 7 target species and "other" fish. "Truth" column values for each of the species are associated with the number of a particular species identified by the video analyst expert, excluding unmarked fish. "Predicted" column values for each of the species represent the number of correctly identified fish for a particular species. "Predicted/Truth" represents the ratio of the Predicted fishes identified over the Truth fishes identified, which is the metric used to assess accuracy of species identifications made by the consensus annotations.

	Truth	Predicted	Predicted/ Truth
Ehu	11	11	1
Opakapaka	25	20	0.800
Kalekale	67	25	0.373
Gindai	12	12	1
Onaga	5	3	0.600
Lehi	41	30	0.731
Hapuupuu	10	8	0.800
Other	56	41	0.732

usability of the annotation data generated from the project. The test image set consisted of 100 images, with the majority of these images containing fish in them. For these fish images, we purposely chose images that contained Deep 7 species, as this was the spread of test images we believed would provide information on how well users would be able to identify each of the target species we were interested in gathering data on. After the beta test was conducted, we realized that the selectivity of the images chosen for the beta test may have caused a bias that affected the resulting beta test data accuracy, which could account for the discrepancy between patterns seen in the beta test data versus the public launch data.

Due to the lessons that we learned through the beta test and public launch of OceanEYEs, there are a couple key things we would recommend to those interested in using citizen science in support of fisheries sciences. First, it is important to provide volunteers supplementary materials to aid them with the tasks you are asking them to perform, and it is especially important for these supplementary materials to contain informative and educational tools that are not extraneous in order to not oversaturate volunteers. Secondly, we recommend that a pilot/beta test be conducted as a means to assess the accuracy of the data produced by the citizen science project. Test sets used in the pilot/beta test should provide information on the data of interest and should also be representative of sets that volunteers will see once the project becomes open to the public.

NEXT STEPS

To ensure a successful citizen science project with sustained participation and solid methods of evaluating progress towards accuracy goals, thinking forward with next steps is paramount. At each stage of our project, we assess the incoming data and identify ways to improve the annotation process. One such method just implemented into the OceanEYEs workflow is a toggle feature, which will assist with finding fish identifications and counts. This feature utilizes a pair of photos that the user can "toggle" between in order to track movement between sequential frames, further improving the accuracy of the data collected. Including movement into the workflow can provide volunteers a way to identify fish far in the distance that may have been missed or allow volunteers to change the fish position to a suitable angle, increasing the possibility of identification.

Another next step planned for the OceanEYEs project is to ingest the annotated OceanEYEs data set images into an automation software developed to identify fish. In order for the artificial intelligence (AI) software to identify our target fish, the model will need to have the pipeline train detectors and classifiers with target fish already tagged in photos. The OceanEYEs citizen science platform is a way to provide the AI with a plethora of identified fish photos. After the AI has proven to provide accurate results, the program can be used for future fish survey missions, allowing for increased deployments with rapid data turnaround times. This increase in operations and subsequent data sets would allow a more comprehensive look into the marine environment, strengthening the stock assessment report.

REFERENCES

Amin, R., B. L. Richards, W. F. X. E. Misa, J. C. Taylor, D. R. Miller, A. Rollo, C. Demarke, H. Singh, G. C. Young, J. Childress, J. E. Ossolinski, R. T. Reardon, and K. H. Koyanagi. 2017. The modular optical underwater survey system. Sensors 17:1–14.

Campbell, M. D., A. G. Pollack, C. T. Gledhill, T. S. Switzer, and D. A. DeVries. 2015. Comparison of relative abundance indices calculated from two methods of generating video count data. Fisheries Research 170:125–133.

Cappo, M., E. S. Harvey, and M. Shortis. 2006. Counting and measuring fish with baited video techniques: an overview. Australian Society for Fish Biology, Hobart, Tasmania.

Cox, J., E. Y. Oh, B. Simmons, C. Lintott, K. Masters, A. Greenhill, G. Graham, and K. Holmes. 2015. Defining and measuring success in online citizen science: a case study of Zooniverse projects. Computing in Science and Engineering 17(4):28–41.

Haight, W. R., D. R. Kobayashi, and K. E. Kawamoto. 1993. Biology and management of deepwater snappers of the Hawaiian Archipelago. Marine Fisheries Review 55:20–27.

Kosmala, M., A. Wiggins, A. Swanson, and B. Simmons. 2016. Assessing data quality in citizen science. Frontiers in Ecology and the Environment 14(10):551–560.

Langseth, B., J. Syslo, A. Yau, M. Kapur, and J. Brodziak. 2018. Stock assessment for the main Hawaiian Islands Deep 7 bottomfish complex in 2018, with catch projections through 2022 *in* . National Oceanic and Atmospheric Administration, Report NMFS-PIFSC-69. Honolulu, Hawaii.

Moffitt, R., D. Kobayashi, and G. DiNardo. 2006. Status of the Hawaiian bottomfish stocks, 2004. Pacific Islands Fisheries Science Center, Administrative Report H-06-01, Honolulu, Hawaii.

Raddick, M. J., G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg. 2010. Galaxy Zoo: exploring the motivations of citizen science volunteers. Astronomy Education Review [online serial] 9(1). Available: https://bit.ly/3ODwsW8 (July 2022).

Richards, B. L., S. G. Smith, J. S. Ault, G. T. DiNardo, D. R. Kobayashi, R. Domokos, J. Anderson, J. Taylor, W. Misa, L. Giuseffi, A. Rollo, D. Merritt, J. C. Drazen, M. E. Clarke, and C. Tam. 2016. Design and implementation of a bottomfish fishery-independent survey in the main Hawaiian Islands. National Oceanic and Atmospheric Administration, Report NMFS-PIFSC-53, Honolulu, Hawaii.

Richards, B. L., O. Beijbom, M. D. Campbell, M. E. Clarke, G. Cutter, M. Dawkins, D. Edgington, D. R. Hart, M. C. Hill, A. Hoogs, D. Kriegman, E. E. Moreland, T. A. Oliver, W. L. Michaels, M. Placentino, A. K. Rollo, C. H. Thompson, F. Wallace, I. D. Williams, and K. Williams. 2019. Automated analysis of underwater imagery: accomplishments, products, and vision. National Oceanic and Atmospheric Administration, Report NMFS-PIFSC-83, Honolulu, Hawaii.https://doi.org/10.25923/0cwf-4714

Schobernd, Z. H., N. M. Bacheler, and P. B. Conn. 2013. Examining the utility of alternative video monitoring metrics for indexing reef fish abundance. Canadian Journal of Fisheries and Aquatic Sciences 71(3):464–471.

Simpson, R., K. R. Page, and D. De Roure. 2014. Zooniverse: observing the world's largest citizen science platform. Association for Computing Machinery, Seoul, Korea.

Western Pacific Regional Fishery Management Council. 2010. Bottomfish fisheries in the Hawaii archipelago, Honolulu, Hawaii. AFS