EPA Public Access

Author manuscript

J Great Lakes Res. Author manuscript; available in PMC 2021 October 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

J Great Lakes Res. 2020 October 1; 46(5): 1469–1478. doi:10.1016/j.jglr.2020.07.009.

Deep Lake Explorer: A web application for crowdsourcing the classification of benthic underwater video from the Laurentian Great Lakes

Molly J. Wick^{a,*}, Ted R. Angradi^b, Matthew B. Pawlowski^c, David Bolgrien^b, Rick Debbout^d, Jonathon Launspach^d, Mari Nord^e

^aORISE (Oak Ridge Institute for Science and Education), Oak Ridge, TN 37831, USA

^bGreat Lakes Toxicology and Ecology Division, Center for Computational Toxicology & Exposure, U.S. Environmental Protection Agency, Office of Research and Development, Duluth, MN 55804, USA

^cGreat Lakes National Program Office, U.S. Environmental Protection Agency, Chicago, IL 60604, USA

^dGDIT, Falls Church, VA 22042, USA

^eU.S. Environmental Protection Agency, Region 5, Chicago, IL 60604, USA

Abstract

Underwater video is increasingly used to study aspects of the Great Lakes benthos including the abundance of round goby and dreissenid mussels. The introduction of these two species have resulted in major ecological shifts in the Great Lakes, but the species and their impacts have heretofore been underassessed due to limitations of monitoring methods. Underwater video (UVID) can “sample” hard bottom sites where grab samplers cannot. Efficient use of UVID data requires affordable and accurate classification and analysis tools. Deep Lake Explorer (DLE) is a web application developed to support crowdsourced classification of UVID collected in the Great Lakes. Volunteers (i.e., the crowd) used DLE to classify 199 videos collected in the Niagara River, Lake Huron, and Lake Ontario for the following attributes: round goby, *Dreissena*, and aquatic vegetation presence, and dominant substrate type. We compared DLE classification results to expert classification of the same videos to evaluate accuracy. Depending on the attribute, DLE had 77% (hard substrate) to 90% (vegetation presence) agreement with expert classification of videos. Detection rates, or the number of videos with an attribute detected by both volunteers and an expert divided by the number where only the expert detected the attribute, ranged from 62% (hard substrate) to 95% (soft substrate) depending on the attribute. Many factors affected accuracy, including video quality in the application, video processing, abundance of species of interest, volunteer experience, and task complexity. Crowdsourcing tools like DLE can increase timeliness and decrease costs but may come with tradeoffs in accuracy and completeness.

*Corresponding author; wick.molly@epa.gov.

Keywords

underwater video; benthos; Great Lakes; crowdsourcing; citizen science; *Dreissena*; round goby

Introduction

The application of underwater videography (UVID) to complement traditional methods for assessing the benthic conditions in the Great Lakes ecosystem has increased in the past decade (Karatayev et al., 2018). UVID can provide data in areas with rocky bottom where grab sampling is difficult or impossible (Lietz et al., 2015; Karatayev et al., 2018). Video is especially useful in Great Lakes connecting channels like the Niagara River, which can have swift currents and extensive areas of rocky substrate (Mehler et al., 2018). Benthic UVID can supply information about native and invasive demersal fishes and invertebrates, substrate, habitat patchiness, plant communities, anthropogenic features, and species interactions and behavior that is generally not obtainable from grab samples.

In the Great Lakes, UVID has been primarily used to assess the status of the invasive round goby (*Neogobius melanostomus*), and zebra (*Dreissena polymorpha*) and quagga mussels (*Dreissena rostriformis bugensis*). Invasions by *Dreissena* and round gobies have profoundly affected Great Lakes benthic ecosystems. *Dreissena* alters physical habitat for other benthic invertebrates and increases sedimentation and mineralization. Their filter feeding affects nutrient cycling, increases water clarity, and is a trophic link between the pelagic and benthic components of the Great Lake ecosystems (e.g., Mayer et al., 2014). *Dreissena* invasion has been associated with shifts in phytoplankton density and community composition, declines in the native amphipod, *Diporeia* spp., and declines in native unionid and sphaeriid bivalves, oligochaetes, and chironomids (Burlakova et al., 2014; Karatayev et al., 2015).

Round gobies can outcompete native demersal fish species for habitat and prey and may consume native species' eggs and fry (Kornis et al., 2012). Gobies consume *Dreissena* (Lederer et al., 2008) and are themselves consumed by piscivorous fish (Madenjian et al., 2011). There is evidence of the transfer of contaminants filtered from water by *Dreissena* from gobies to higher trophic levels (Johnson et al., 2005). Unfortunately, both round goby and *Dreissena* are often not vulnerable to conventional sampling methods (e.g., trawling and PONAR, respectively). Researchers have used UVID to quantify *Dreissena* coverage, density, and biomass (Ozersky et al., 2011; Karatayev et al., 2018), and as an early detection monitoring tool for invasive species (Trebitz et al., 2019).

Underwater video is an affordable and efficient means of sampling and assessing benthic species and habitats across the Great Lakes. Cameras and associated gear are relatively inexpensive (e.g., compared to divers) and require little maintenance, and video can be collected in under five minutes by lowering a camera and lights mounted on a carriage to the bottom. UVID was tested as an assessment tool in the Great Lakes nearshore as part of the EPA's National Coastal Condition Assessment in 2010 (Lietz et al., 2015). Video paired with grab sampling improved detection of *Dreissena* over grab sampling alone. Even so, as many as 45% of the videos collected had poor or marginal image quality. Fortunately,

improvements in cameras, gear, and sampling methods, have increased the quality of UVID collected in lake-wide benthic assessments since 2010.

Although video quality, defined broadly as features of video footage that affect interpretation (e.g., resolution, brightness/darkness, contrast, color accuracy, focus, depth of field, content), has improved over early efforts, video classification remains challenging (see videos 1, 2, 3, 4 and 5 in Appendix 1 for examples). Consistent and accurate video interpretation is difficult because of variation in water clarity, ambient lighting, bottom relief, amount of current, and movement of organisms (Salman et al., 2016). Like most sampling methods, detecting organisms or objects at low abundance is difficult (e.g., new *Dreissena* invasions, Trebitz et al., 2019). Where invaders are still rare, there is a low probability of a video being collected at occupied sites, and a high probability that they will be overlooked, especially in visually complex habitats. To address these challenges and capitalize on the efficiency and affordability of UVID as an assessment tool, extensive data sets, and affordable and accurate UVID classification tools are essential.

By engaging the public, crowdsourcing video classification has potential to cost-effectively address these challenges (e.g., Swanson et al., 2016; Happel et al., 2020). Crowdsourcing allows multiple volunteers to classify each UVID, resulting in replicate classifications and consensus estimates (e.g., Good and Su, 2013; Swanson et al., 2016). Ideally, crowdsourcing can increase timeliness of reporting, reduce classification costs, and facilitate accurate analysis of large datasets.

We developed a web application, the Deep Lake Explorer (DLE), to explore the utility of crowdsourcing for classification of UVID collected in the Great Lakes. In this paper, we describe the development and implementation of DLE, evaluate its accuracy, and discuss its limitations and potential as a tool for UVID classification.

Methods

Video collection

Underwater video was collected in 2017–2018 in the Niagara River, Lake Huron, and Lake Ontario. We collected UVID in the Niagara River (Fig. 1) as part of a National Coastal Condition Assessment pilot study (NCCA) of Great Lakes connecting channels (Pawlowski et al., in prep). In Lake Huron (2017) and in Lake Ontario (2018, Fig. 1), we collected UVID as part of the Cooperative Science and Monitoring Initiative (CSMI, ECCC and USEPA, 2018), lakewide assessments in the Great Lakes coordinated by USEPA. Video survey and collection methods are described in detail in Appendix 2.

Two cameras (GoPro Hero5s, San Mateo, California; or SeaViewer 6000 SeaDrop cameras, Tampa, Florida) and lights were mounted on a custom-built stainless-steel carriage, which could be lowered to the lake bottom (Fig. 2). One camera and one or two lights were each mounted in the down-looking and the oblique-looking direction. In this study, we only used oblique video because it was superior to down-looking video for detecting and identifying demersal fish (see Appendix 1 videos 5 and 6 for examples of oblique and down-looking footage, respectively). The camera carriage rested on the bottom for approximately 1 minute

and was then either raised to the ship, or raised off the bottom while the boat drifted, and then dropped to the bottom again for one minute for another “drop”. Multiple drops were only collected in Lake Ontario and were classified separately. In the Niagara River, PONAR failures due to rock substrate were common (16 sites) and additional video was collected at substitute off-channel locations with soft substrates. Lake Huron video was collected at up to 24 random locations >5 m deep and within 2 km of the base sites shown in Fig. 1 (mean depth: 12 m, maximum: 48 m). Depth of stations in Lake Ontario were 11 – 36 m (mean: 21 m), and depth of stations in the Niagara River were 1 – 20 m (mean: 7 m). Table 1 shows the number of sites and videos collected and classified in each waterbody.

Video processing

To minimize the amount of raw UVID footage needing review, we discarded segments of videos that were irrelevant (descent to bottom, ascent back to surface) or were otherwise unusable (poor visibility due to suspended sediment) using ShotCut software (MeltysTech, LLC, San Francisco, California). However, the 2–3 seconds of video before the video carriage reached the bottom and after the start of its ascent to the surface were included for the wider area of the bottom visible from the suspended camera. We determined that classification of UVID was more reliable than screen captures (stills) for several reasons. Randomly-selected screen captures may be of poor quality due to transient suspended sediment. Substrate complexity often makes fish difficult to detect if they are not moving. Lastly, moving objects (e.g., fish, vegetation waving in current) may be out of focus in screen captures if camera settings (ISO, shutter speed) are not optimal for benthic light conditions.

Trimmed videos were segmented to 15 second clips to make classifications simpler and quicker and to retain volunteers’ attention (Eveleigh et al., 2014). We segmented videos using ffmpeg (FFmpeg Developers, 2016) and the Python packages imageio, moviepy, panoptes-cli, and pandas (Zulko, 2017; Klein et al., 2019; Reback et al., 2019). Clips at the end of videos that were > 5 seconds and < 15 seconds were included in DLE. Some full videos were very short (< 20 seconds) due to the camera carriage tipping over or high current on the Niagara River. These videos still supply useful information and were included if 2 seconds or longer.

Videos were compressed and reduced from their original high definition 1920×1080 -pixel resolution to standard definition 960×540 -pixel resolution before uploading to the Zooniverse platform to reduce file sizes to about 1MB and ensure rapid loading time with typical internet speeds. A total of 746 video clips were uploaded to DLE for classification. Scripts to process videos are available at https://github.com/debboutr/to_the_zooniverse.

Deep Lake Explorer

DLE was developed for and hosted on the Zooniverse platform (<https://www.zooniverse.org/projects/usepa/deep-lake-explorer>), which supports online citizen science projects in which volunteers participate in research by helping to classify visual or audio information (Zooniverse, 2019). Most Zooniverse projects classify still images; DLE is one of few projects created to classify video footage. Few other studies (e.g., Di Salvo et al., 2013;

Mugar et al., 2014; Matabos et al., 2017; Ferrini et al., 2018), and none in the Great Lakes, have used crowdsourcing to classify UVID footage.

To increase the relevance of project outcomes for management, an advisory group of scientists and natural resource managers from all Great Lakes states and several federal agencies was convened to provide input on DLE. The group advised us to focus on classification of benthic invasive species, especially round gobies and *Dreissena* mussels, and their associated habitats. We developed and revised classification “workflow” questions (Table 2) based on input from the advisory group and from DLE beta testers (described below). A DLE tutorial, field guide, frequently asked questions, related links, example videos, and additional background information could be accessed from the workflow as resources for volunteers (available at <https://www.zooniverse.org/projects/usepa/deep-lake-explorer>). DLE included a user forum through which volunteers could comment or ask researchers questions about video clips. Video 7 in Appendix 1 shows the DLE tutorial and workflow.

Fifty video clips were beta tested by a subset of DLE volunteers. We selected clips that represented the range of conditions present in the Great Lakes, including fine and coarse sediment, low and high current velocity, and vegetated and unvegetated bottom. Most selected clips depicted mussels, round gobies, or both. One hundred fifty-nine volunteers participated in beta testing in April 2018. Based on the beta test results and volunteer feedback, we made several changes to the workflow (Table 2), including simplifying the questions to binary responses and adding questions about relative abundance of mussels and observations of human-made objects (neither discussed further here).

DLE was launched and added to the Zooniverse homepage in June 2019. Notices requesting participation in the project were circulated among advisory group networks. Each clip was classified by ten volunteers, after which the clip was retired such that no more volunteers saw it. The requirement of ten users was selected based on beta testing and Zooniverse recommendations. Within two weeks, every clip had been reviewed by ten volunteers. During the project, the first author moderated the user forum and responded to volunteers’ questions and comments.

Expert classification

Each full video (not each clip) was classified by an expert for *Dreissena* and round goby presence, vegetation presence, and substrate type. See Appendix 3 for complete expert analysis methods. In brief, two scientists, authors TRA and MJW, both experienced in identification of *Dreissena* and round goby from videos, first independently reviewed Lake Ontario videos. Differences between the two experts’ classifications were reconciled by reviewing the video together for a final classification for Lake Ontario, and the Lake Huron and Niagara River videos were classified by TRA with support from MJW.

Data analysis

The volunteer classifications for each clip were exported from DLE in JavaScript object notation (JSON) format and analyzed in R (R Core Team, 2019) using the tidyjson package (Stanley, 2017). For *Dreissena* presence, vegetation presence, and substrate type (dominantly

hard or dominantly soft, Table 2), we first aggregated all volunteer classifications for all clips from each video. Results were pooled because these fixed or sessile attributes do not change during a video, so all clips should yield the same result. We then defined a threshold in percent volunteer agreement, above which vegetation and *Dreissena* were considered present and substrate was classified as “hard” for the entire video.

For round goby, we used a slightly different method because round gobies are active and may swim in and out of a video, so each clip may not return the same result. The volunteer percent agreement threshold was applied for each *clip*’s classifications (rather than each video’s classifications), above which round goby were considered present in the clip. If round gobies were considered present in any clip within a video, then they were present in the video by default. Video quality can also vary from clip to clip due to changes in turbidity, so this method was also used for the video quality attribute (Table 2).

For round goby presence, *Dreissena* presence, substrate type, and vegetation presence, we selected volunteer agreement thresholds based on the threshold with the highest accuracy, and lowest rate of false negatives. We defined accuracy as the percent of videos (aggregated from clips) for which the DLE and expert video classification agreed. False negatives were instances when an attribute was detected by the expert but was not detected by DLE with the volunteer agreement threshold applied. We also considered false positives, instances when DLE volunteers detected an attribute that experts did not observe. The videos represent a broad range of benthic conditions in the Great Lakes, so we pooled all the data to identify volunteer agreement thresholds applicable in all three waterbodies.

To determine if the removal of poor-quality clips from the dataset increased DLE accuracy, we compared video classification of the full dataset to a dataset with poor-quality videos removed. We defined poor-quality videos as those for which 70% of volunteers agreed the quality of the video was inadequate to perform the classification. For round goby, *Dreissena* presence, vegetation presence, and substrate type (dominantly hard or soft), we calculated DLE detection rates as the number of videos with the attribute detected by both DLE and expert divided by the number of videos where the expert detected the attribute. We compared DLE and expert classification results for each waterbody.

Results

Participation in Deep Lake Explorer

Within two weeks of launching, 531 volunteers classified all 746 UVID clips in DLE. Ten volunteers classified more than 100 clips each; 61 volunteers classified 21 to 100 clips each, and 460 volunteers classified 20 or fewer clips each. Six percent of volunteers commented in the forum and left 192 comments or questions. Comments included observations about video content (61%), questions about the project and the workflow (13%), troubleshooting questions or comments (7%), suggestions for the project (3%), and responses to other comments (17%).

Deep Lake Explorer volunteer agreement thresholds

For higher volunteer agreement thresholds, DLE-expert agreement was lower, ranging from 45% (*Dreissena* presence) to 69% (vegetation). If a higher threshold was applied, the DLE-expert agreement declined (Fig. 3) because the occurrence of false negatives (when an attribute was detected by the expert but not by DLE) increases. Attributes can go undetected by volunteers for several reasons, such as if a round goby is only visible for a few seconds, or because the round goby, *Dreissena*, or vegetation is small, motionless (goby) and/or cryptic (Appendix 1: video 8). All false negatives for round goby presence were due to volunteers missing fish in the clip rather than from volunteers observing a fish but not identifying it as a round goby. False negatives were most common for *Dreissena* presence, likely because *Dreissena* can be cryptic, requiring an experienced eye to detect. Mussels can be especially hard to detect if they are not present in the foreground or if they are covered by algae (Appendix 1: video 9).

False positives, or instances when DLE volunteers detected an attribute that experts did not, were less common. There were a few clips where a volunteer misidentified another species of fish or a moving piece of algae or vegetation as a round goby. Applying a volunteer agreement threshold of 30%, there were no occasions when DLE volunteers identified a round goby that expert classification missed (Fig. 3). False positives were most prevalent at low thresholds and were the highest for vegetation.

Based on Fig. 3, we selected the thresholds presented in Table 2 for classification of videos. For *Dreissena* and vegetation presence, DLE-expert agreement was the highest with a volunteer agreement threshold of 40% or less (Table 3, Fig. 3). We selected a threshold of 40% to minimize false positives. For round goby presence, a lower threshold of 30% volunteer agreement resulted in the peak DLE-expert agreement of 90%. The greatest DLE-expert agreement for substrate was at 50% volunteer agreement, partially because there were two “present” options (dominantly soft or hard) rather than presence/absence options.

Video classification accuracy and uncertainty

The agreement between two expert reviewers, based on classification of videos from Lake Ontario, ranged from 90% (substrate type) to 96% agreement (round goby presence, Table 3). DLE accuracy, based on agreement between DLE and the final expert classification for the entire dataset, ranged from 77% (substrate type) to ~90% (round goby and vegetation presence, Table 3). With the thresholds we selected, DLE results underestimated presence of round gobies, hard substrate, and *Dreissena*, and overestimated vegetation presence (Table 3).

The detection rate (number of videos with attribute detected by both DLE and expert divided by the number of videos where expert detected the attribute) for vegetation was the highest of all attributes at 93% and was in fact overestimated (Fig. 2), likely because of the semi-subjective nature of its definition: “taller than a surface-coating or fuzz.” The 78% detection rate for round gobies was higher than for *Dreissena* (66%, Table 3). *Dreissena* are sessile and can be obscured by other objects, whereas the eye is naturally drawn to the movement of a fish. The detection rate for soft substrate was higher (95%) than for hard substrate (62%).

The difference in the percent of videos with round gobies detected by experts and by DLE volunteers was highest for Lake Ontario and lowest for Lake Huron (Table 4). Lake Huron videos were taken during daytime when round gobies are most active and probably easier to detect; many Lake Ontario videos were collected at night. Conversely, the difference in the percent of videos with *Dreissena* detected by experts and by DLE was largest for Lake Huron, and smallest for Lake Ontario. Fewer mussels, more complex substrates, and slightly poorer video quality from Lake Huron is likely responsible for this discrepancy. The difference in the percent of videos with hard substrate detected by experts and by DLE was largest for the Niagara River (Table 4). Niagara River has complex substrates (gravel, sand, and shell hash, which are often mixed and partially obscured by vegetation) that were difficult for volunteers to classify using the dominantly hard or dominantly soft choices in the DLE workflow.

Twenty-nine of the 746 clips (4%) met the threshold of 70% volunteer agreement that video quality was poor. Experts verified these videos as being of poor quality. Removing these clips changed DLE agreement with expert classification by 1% for round goby presence, substrate type, mussel presence, and vegetation presence. Percent volunteer agreement was expected to be the lowest for poor quality videos and videos with classifications disagreeing with experts. For vegetation and substrate type, mean percent volunteer agreement for all clips was lower for videos where the DLE and expert result disagreed than for videos where the DLE and expert result agreed (Fig. 4). However, for round goby presence and *Dreissena* presence, mean percent volunteer agreement for false negatives was higher than or similar to true positives.

Discussion

Selecting appropriate volunteer agreement thresholds

We used volunteer agreement thresholds based on agreement of DLE classification with experts to determine the result for each clip or video. The videos classified in DLE represent the range of benthic conditions within the larger Great Lakes, so the thresholds used here could be applied for future classification of similar attributes in Great Lakes videos. However, the variation in thresholds among attributes (e.g., 30% for round gobies; 50% for substrate type) suggests that other attributes will require independent verification using an expert-classified dataset to identify appropriate volunteer agreement thresholds. Project objectives should be considered when selecting appropriate thresholds as some projects may prioritize reducing false negatives over false positives (Torre et al., 2019).

Classification accuracy

For all attributes, DLE did not perform as well as experts at classification, with accuracy rates lower than the agreement between two experts' classifications. Substrate had the lowest agreement between two experts (90%) of all attributes, illustrating that making the distinction between hard and soft is surprisingly challenging. Niagara River had the most videos with substrate classified differently in DLE and by the expert (32/86 in Niagara; 45/196 total). Based on DLE, hard substrate detection rate was 62% and soft substrate detection rate was 95%, revealing a classification bias in volunteers towards soft substrate.

This likely resulted from users following directions given in the tutorial to select soft sediment when a sediment cloud was observed or when vegetation obscured the bottom. While these instructions were generally reliable, there were instances when dense vegetation (e.g., filamentous algae) nearly obscured boulders, or where fine sediment coated bedrock and led to an incorrect classification. Sites where the bottom was completely obscured by *Dreissena* were also challenging to classify. Some clips contained both fines and coarse materials in similar proportions, making selection of a dominant substrate type difficult. For workflow simplicity, we did not include an “equally dominant” option, because it is subjective and challenging to define and apply. We only included oblique videos in this study, but the use of down-looking video footage (collected simultaneously) along with oblique footage to classify substrate at a site may reduce error for this attribute. Down-looking footage offers direct observation of the substrate (and a scale for measurement of particles), and oblique footage gives a broader perspective of the larger area and substrate patchiness.

For round goby presence, *Dreissena* presence, and vegetation presence, DLE also had accuracy rates lower than the agreement between two experts. Several factors that likely affected the accuracy of attributes in DLE were video segmentation methods, video resolution and video quality, abundance of object or species of interest, object or species movement and time at bottom, and volunteers’ level of experience.

Video segmentation and trimming affected DLE accuracy. We removed portions of videos that did not occur on the bottom or had excessive turbidity because we wanted to determine if volunteers could classify typical UVID collected in the Great Lakes. In some cases, experts made observations of organisms or substrate type only visible or discernable before the carriage landed or after it was raised off the bottom, portions of the video which were not included in DLE clips. To be accurate at the site scale, those observations should be included. Because those portions of the videos were omitted from DLE clips, they reduced accuracy at the site scale by increasing false negative detections. Likewise, if the last clip of a video was shorter than 5 seconds, it was not included in DLE, although a round goby may have only been present or visible during that time. To both address this problem and eliminate the need for manual video processing, all UVID footage could be classified as part of the workflow in DLE. A multi-stage workflow could be employed in which volunteers first classify all clips as useful (the clip shows the lake bottom) or not useful (clip does not show bottom), then classify useful clips for the attributes of interest.

Video resolution and video quality can also affect DLE accuracy. Video resolution refers to the number of horizontal and vertical pixels in the imagery, and video quality refers more broadly to features of imagery that affect interpretation including resolution, brightness/darkness, contrast, color accuracy, focus, depth of field, as well as ambiguous content. Ambiguity can stem from site complexity; for example, when objects obscure attributes. Videos on DLE were lower resolution than the versions classified by experts (see Appendix 1: videos 10 and 11). This issue will persist for video-based projects until typical internet speeds increase to allow rapid streaming of full-resolution, uncompressed video files. An early version of DLE was tested using videos collected in NCCA in 2010 and 2015 with cameras lowered to the bottom on a weighted line (Lietz et al., 2015, Appendix 2). Many

videos were not of sufficient quality for volunteers to classify successfully, and beta testers expressed frustration with poor video quality. Improvements in both camera resolution and deployment methodology, especially the use of a video carriage that rests on the bottom, resulted in the higher-quality videos used in this study. However, some videos were still challenging to classify due to low contrast or poor focus, combined with ambiguity due to site complexity.

Abundance of the object or species at the site and in the waterbody can affect accuracy and detection rates. Swanson et al. (2016) found rare species had higher rates of both false positives and false negatives, but Crall et al. (2011) found false negatives were more common than false positives in invasive species detection by volunteers. Based on volunteer feedback, low abundance of target species may cause volunteers to feel like they are not contributing much to science by their efforts. However, Bowyer et al. (2015) found that participation time increased as the number of “blanks”, or subjects without the object of interest, increased. Intermittent reward theory suggests volunteers may be willing to spend more time volunteering in hopes of finding what they are directed to look for (Kendall, 1974). So, for rare attributes or organisms, increasing the number of reviews required to increase the probability of volunteers seeing it may not incur a significant cost in terms of additional volunteers needed, if volunteers watch more clips in hopes of finding a rare organism.

Movement and the amount of time present in the video also likely affects accuracy of round goby observations. The Niagara River had a lower abundance of round gobies than Lake Huron or Lake Ontario (Table 4). Videography in the river was difficult due to swift currents. Short bottom times may have affected detection rates. The mean video time on bottom for Niagara River videos in which gobies were detected was 26 seconds longer than videos in which gobies were not detected, a significant difference ($P < 0.05$, US EPA unpublished data). For both round gobies and *Dreissena*, false negatives were much more common than false positives (Fig. 4), so our results represent minimums and our detection rates can be used to estimate percent presence.

Volunteers' level of experience likely affects their ability to detect species present, especially at low density (Fitzpatrick et al., 2009; Kelling et al., 2015). Some crowdsourcing projects, especially those which compensate users for their classifications like those hosted on Amazon Mechanical Turk (<https://www.mturk.com/>), have attempted to weight or flag data for additional review based on the level of user experience (e.g., Rahman et al., 2015). However, this complicates data analysis because methods for defining adequate experience must be determined. This may also be inappropriate for citizen science projects because it ignores some volunteers' contributions, and ultimately requires more volunteers and/or experts to fulfill data needs.

Perceptions of Video Quality

Removal of clips identified as inferior-quality by volunteers (70% agreement among volunteers that quality was poor) from the dataset did not improve DLE-expert agreement. This is not surprising because even if a single clip had poor quality due to a transient sediment cloud, most videos had other clips with adequate quality to classify. We expected

poor quality videos to have low volunteer agreement for the other attributes, but that was not the case. The intention of the video quality question was to determine video quality for all attributes, but video quality may be different with respect to each attribute. Even if a volunteer thought the video was poor quality, they were first required to respond to all other questions. For presence/absence questions, volunteers usually chose “not present” even if they thought video quality was too poor to make a reliable determination. For this reason, volunteer agreement for false negatives was higher than that for true positives (Fig. 4).

Other projects have used measures of volunteer agreement as a tool to gauge uncertainty or difficulty of a subject (Swanson et al., 2016; Jiménez et al., 2018; Torre et al., 2019). This does not work if participants can choose absent if a video is of poor quality. Some crowdsourcing projects include an “I don’t know” option (Jiménez et al., 2018; Torre et al., 2019). Subjects with high agreement for “I don’t know” can be eliminated or flagged for additional review, and uncertainty can be quantified for the remaining subjects based on their degree of agreement. We did not include this option because it can result in volunteers avoiding classification of difficult subjects, which increases the number of volunteers needed to complete a project (Swanson et al., 2016; Salk et al., 2017). However, its inclusion would reduce the misclassification of “not present” for videos that are too poor to assess and could also improve substrate assessment when the bottom is obscured by vegetation or mussels.

Implications for project design

We modeled DLE on other successful crowdsourcing projects with an emphasis on ease of use, an aesthetically pleasing interface, and facilitating exchange between volunteers and scientists (e.g., Raddick et al., 2009; Dickinson et al., 2012; Rotman et al., 2012; Eveleigh et al., 2014; Nov et al., 2014). Other crowdsourcing projects have achieved higher accuracy than ours, but these were based on simpler tasks: classifying relatively unambiguous objects in still images (e.g., terrestrial megafauna, Swanson et al., 2016; or pollution threats, Torre et al., 2019). Simplicity is essential to keep volunteers engaged (Eveleigh et al., 2014), but may entail a tradeoff in accuracy. Often the videos were not simple or uniform, and the tutorial and other guidance could not describe all situations presented in the videos such as *Dreissena* with attached algae on distant boulders (Appendix 1: video 8) or motionless round gobies half-hidden by mussels (Appendix 1: video 9). Providing the right amount of information and guidance to volunteers is critical to increase the chance that each new volunteer makes it through the first clip without getting discouraged or bored. Ideally, a tutorial provides just enough information to successfully classify most videos and prompts the volunteer to refer to the field guide if they encounter something not described in the tutorial.

Keeping volunteers engaged over time and preventing users from exiting the application is a common challenge for crowdsourcing projects. Like other Zooniverse projects (Eveleigh et al., 2014), most DLE participants classified only a few clips (median = 5 clips; mean = 14 clips). A forum where volunteers can interact with project moderators and with other volunteers can help engage and retain volunteers (Nov et al., 2014). One active volunteer, who classified 359 clips, commented that “This was by far the most fun I’ve had on a Zooniverse project, and it’s in large part due to [the moderator’s] interaction with the

volunteers in the Talk section.” While most volunteers did not post to the forum, volunteers can learn how to classify clips and improve their skills by reviewing others’ comments and questions (Mugar et al., 2014). For larger datasets, asking experienced volunteers to take on additional responsibilities like helping moderate the forum can be rewarding, magnify their contribution to the project, and can help keep them engaged (Nov et al., 2014). It can also help alleviate the moderation burden for scientists for projects with large datasets.

Authors have identified tradeoffs in cost, timeliness, and accuracy of results between traditional science and citizen science (Gardiner et al., 2012; Garcia-Molina et al., 2016). Classification complexity and the scientific utility of results may be limited by the reliance on non-expert volunteers because relatively simple workflows are necessary to engage and retain volunteers. For DLE, more specific and well-defined substrate questions and additional guidance could reduce the effect of this tradeoff. More complex tasks could be incorporated in DLE by breaking them into multiple simple steps or into separate workflows with additional related training and support materials. However, additional workflows may increase the total project time and moderation burden.

The educational value of crowdsourcing projects could potentially justify some loss of scientific complexity if a simpler workflow promotes participation. For some projects, public education may be an inherent project priority, or may be required as a proviso of funding. Indeed, a core objective of Zooniverse is “to engage with the public in order to educate and change attitudes towards science” (Cox et al., 2015, p.1). Several studies have shown that participants in citizen-science programs demonstrate greater scientific knowledge, skills, and positive attitudes toward science and the environment than the general public (Brossard et al., 2005; Bell et al., 2008; Bonney et al., 2009; Masters et al., 2016).

The utility of crowdsourcing for video classification

Crowdsourcing has the potential to bridge the gap to the next generation of video analysis tools: machine-learning models (Salman et al., 2016). In the long-term, machine learning will accelerate and improve classification of large visual datasets (Kamar et al., 2012), but current model applications are inefficient in complex benthic habitats (Matabos et al., 2017) like those in the Great Lakes. To train models in these habitats, large annotated video datasets are needed, which DLE could help create.

In the near-term, DLE can help meet managers’ needs for improving assessment of benthic conditions in the Great Lakes. Here, we focused on detecting the presence of two invasive species relevant for Great Lakes component of the National Coastal Condition Assessment. The NCCA collects significant underwater video data, approximately 2000 minutes or 33 hours of video every 5 years. For smaller datasets, expert classification may be a more cost-effective approach to classification, unless public engagement and education is an explicit program goal. Assuming each video takes approximately 10 minutes to analyze, it takes ~330 hours to classify them all. In DLE, this equates to 7,800 15-second clips, which is a relatively small dataset compared to other Zooniverse crowdsourcing projects. DLE requires up-front effort in terms of design, and ongoing involvement for oversight and moderation. We found, however, that after the first week, experienced volunteer classifiers began answering new volunteers’ questions, reducing moderation requirements for scientists.

Crowdsourcing is not as accurate as expert analysis, but invasive species detection rates from DLE could be used to estimate percent presence for future analyses of the same or similar attributes.

Developing sampling and assessment methods that are comparable across Great Lakes benthic habitats (e.g., shallow and deep; soft and hard sediments) is a challenge for scientists and managers. Experience from seafloor observatory networks has shown that manual processing of huge underwater video datasets is very resource intensive, which may be a barrier to establishing similar networks in the Great Lakes (Matabos et al., 2017). Crowdsourcing is a method to classify large video datasets, and could be integrated into a collaborative, international video monitoring network to assess benthic conditions in the Great Lakes at a scale and density previously limited by classification methods available. Collection of high-resolution benthic video is now affordable and adaptable across sample platforms (Appendix 2). The combination of underwater video collection and crowdsourced classification could be applied by non-government organizations, local agencies, and individual citizen scientists for application in the Great Lakes or other inland waterbodies (Locke et al., 2019). This monitoring network could integrate existing benthic assessment efforts focused on detection of invasive species (*Dreissena*, round gobies, the macroalga *Cladophora*, and Asian carp, and other invasive species), but could also include other metrics (e.g., percent cover, number of individuals), and could include imagery collected using alternative collection methods such as down-looking video, video collected via sled, autonomous underwater vehicles, or remotely operated vehicles, or sediment profile imagery.

Acknowledgements

We thank Knut Mehler, Jill Scharold, Dustin Woodruff, and the captains and crews of the US EPA R/V *Lake Explorer II*, *Lake Guardian* and *Mudpuppy II*, who all assisted in data collection; the participants in the DLE advisory group, and Paul Skawinski and Rachael Mady who provided reviews of the manuscript. We thank all volunteers who participated in Deep Lake Explorer, especially SirUrielPerpetua, Tomburgerpie, and an anonymous user who each classified over 350 clips. EPA Office of Research and Development Innovation Office and Great Lakes Restoration Initiative provided funding for components of this project. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. Mention of trade names does not constitute endorsement or recommendation for use.

Appendix 1. Table of video examples.

All examples are full videos (not clips), unless otherwise stated.

Video number	Description	Notes	Site	Latitude	Longitude	Waterbody	Depth (m)	YouTube link
1	Video quality increased significantly since NCCA 2010 & 2015	Example of typical video collected in 2010. Files collected in 2010 were .avi files, and some	NCCAGL10-1193	43.8034	-76.2518	Lake Ontario	12.7	https://youtu.be/yTfRmsWwd1w

		compression occurred converting to .mp4 for posting here. <i>Dreissena</i> are obvious on the rocks but individual organisms are very hard to distinguish. Round gobies are present but are often extremely cryptic (e.g., 0:11, 0:52), although one is obvious swimming at the bottom of the screen at 0:42. Objects not in the foreground are out of focus. No sediment grab sample could be collected at this site.						
2	Video quality increased significantly since NCCA 2010 & 2015	Example of typical video collected in 2015. Initially, the camera is too far from the bottom to make observations. Once it is lowered, <i>Dreissena</i> are obvious. A very cryptic goby can be observed at 1:23, 1:34. The camera movement is often distracting (like shaking and spinning in this example) and at times can be too fast to see the bottom clearly, making footage unusable and potentially could cause motion sickness in viewers.	GLNS15-2190	43.32406	-76.8066	Lake Ontario	7.8	https://youtu.be/_IDHWjsRLF8

3	Video quality increased significantly since NCCA 2010 & 2015	Example of a typical good quality video collected during in the Lake Huron CSMI. The view is well-lit and individual mussels are visible on rocks. Individual round gobies and their dorsal spots are visible swimming among rocks throughout the duration of the video, although individuals are occasionally hidden behind rocks. The characteristic black dorsal spot is visible on the goby in the foreground at 1:35.	Harb05_12_camB	43.78512	-82.60068	Lake Huron	9.5	https://youtu.be/D1QdmtWyUCA
4	Video quality increased significantly since NCCA 2010 & 2015	Example of a typical good quality video collected during the Lake Ontario CSMI. Round goby is present when camera lands on bed of <i>Dreissena</i> , and additional round gobies swim into view. Individual mussels are obvious. Video was collected during nighttime, and round gobies are less active.	DD20_B	43.36455	-77.83355	Lake Ontario	15.5	https://youtu.be/AiwEMG2oMJA
5	Video quality increased significantly since NCCA 2010 & 2015	Typical example of a good quality video collected in the Niagara River. Note the rapid current and brief plume	NIAGT-102_B_01	42.95202	-78.91214	Niagara River	4.2	https://youtu.be/n0luzO7urNs

		of fine sediment after camera frame contacts bottom. Round gobies with obvious dorsal spots appear from behind vegetation. Note the very tiny round goby swimming in lower center of image at 1:00 – 1:06. While this size goby is cryptic with current video methods, previous UVID methods would have missed them.						
6	Example of down-looking video	Example of video collected with down-looking camera. Yellow and black squares on carriage feet each measure 2.54 cm. Note the round goby which darts away quickly as carriage lands.	NIAG-073_A_01	43.02577	-79.05012	Niagara River	4.2	https://youtu.be/pOL34ebnq2c
7	Video walk-through of Deep Lake Explorer	This video shows the process of going through the DLE workflow. The tutorial pops up automatically on browsers that have not visited the page before.	NA	NA	NA	NA	NA	https://youtu.be/4ktk089UuVk
8	Cryptic round goby	This is an example of a video in which DLE volunteers did not identify round goby, but expert classification did. A round goby head is	DD15_B	43.55172	-76.28582	Lake Ontario	18	https://youtu.be/7F1d_2gmUm0

		visible on the left side of the screen until 2:17 when it swims behind mussels. The goby blends into its surroundings well, although the pattern is familiar to a trained eye.						
9	Cryptic <i>Dreissena</i>	This is an example where DLE volunteers did not identify live <i>Dreissena</i> present, but expert classification did. Mussels are cryptic on rocks in the distance, and two mussels are present in left foreground, identifiable by their dark slits despite being covered with algae and vegetation. They are also visible as the camera is lifted off the bottom.	NIAG-005_B	43.06340	-78.96191	Niagara River	2.9	https://youtu.be/LdmVt0GV-M
10	Comparison of original vs DLE video quality	This is an example of a 15 second clip included on DLE, at its original resolution of 1920 × 1080p.	NIAG-017_B	42.99691	-79.02560	Niagara River	7.6	https://youtu.be/V14o0u2lrzs
11	Comparison of original vs DLE video quality	This is an example of a 15 second clip included on DLE, compressed to 960 × 540p resolution.	NIAG-017_B	42.99691	-79.02560	Niagara River	7.6	https://youtu.be/iff6h6St1yzI

Appendix 2. Video collection methods

This appendix describes the methods used to collect the underwater video included in DLE. Our basic approach was to lower a custom-made welded stainless-steel camera carriage to

the lake bottom. Affixed to the carriage were two cameras, one down-looking, and one oblique-looking, and lights (see Table A2.1 for technical details). The carriage was allowed to rest on the bottom until enough video footage was captured or the movement of the ship or boat lifted the carriage off the bottom. At some locations on Lake Ontario, the camera was raised partway to the surface, while the ship was repositioned (or was moved by wind and current) at a site and the camera carriage was re-lowered to the bottom to maximize acquisition of video while on station. On the Niagara River, PONAR failures due to rock substrate were common and additional video was collected at substitute off-channel sites with soft substrates.

Videos were collected during three different research efforts associated with ongoing research programs. Lake Huron was surveyed during the August 2017 Cooperative Science and Monitoring Initiative (CSMI, Great Lakes Sea Grant, 2019) cruise by the USEPA R/V Lake Explorer II. The Niagara River (including Buffalo Harbor) was surveyed in July of 2018 by small boat (6.5 m) as part of a Great Lakes connecting channel assessment (Pawlowski et al., in prep). Lake Ontario was surveyed during the September 2018 CSMI cruise by the USEPA R/V Lake Guardian. The UVID collection locations were determined by the program objectives of each survey effort. For Lake Huron, the survey design was a consensus among researchers on the locations of transects optimized to characterize spatial and temporal variation in lower trophic levels and nutrients across the lake (David Bunnell, US Geological Survey, personal communication). We extended each transect inshore to the 5 m depth contour and collected UVID at up to 12 random locations within 2 km of the transect-based stations and >5 m deep. We successfully collected UVID at 71 locations at 8 stations in Lake Huron, two in the North Channel, three in Georgian Bay, and three in the main basin (Fig. 1). For the Niagara River, we selected sites using a stratified random survey design (USEPA, 2018; Kincaid et al., 2019; Stevens et al., 1999). We successfully collected UVID at 88 locations in the Niagara River. For Lake Ontario, we successfully collected UVID at 52 survey locations (Fig. 1) selected based on historical and ongoing monitoring of the nearshore benthos. The depth, time on bottom, and coordinates of every video was recorded in the field.

Table A2.1

Settings and technical details of UVID systems using in benthic surveys conducted by US EPA and discussed in this paper.

Number	System (application)	Camera system	Deployment method	Lighting	Camera settings	Other details
1	SeaViewer® v1 (Great Lakes NCCA before 2016)	One SeaViewer® 950 SeaDrop color video camera linked via cable to a topside display and recorder	Weighted drop line; oblique view	One SeaViewer® Sea-Light LED light	Default	Height of camera above substrate not controlled. This method was used in Lietz et al. (2015) and in NCCA 2015, and these videos were tested in an early beta test version of DLE.
2	SeaViewer® v2 (Lake	Two SeaViewer® 6000	Custom video carriage - large; down-	Single SeaViewer® Model 10-	30 frames/second, 1080-pixel	Video camera carriage is described in Angradi (2018).

Number	System (application)	Camera system	Deployment method	Lighting	Camera settings	Other details
	Huron 2017 CSMI)	SeaDrop color video cameras linked via cable to topside displays and DVR recorders	looking and oblique view	LED light per camera	resolution, otherwise default settings.	Height of lens of down-looking camera above the substrate = 62 cm. Field of view (i.e., videoquadrat area) of down-looking camera $\approx 35 \times 62$ cm. Useful field of view of oblique camera variable depending on light and obstacles. Vertical height of oblique camera lens above substrate = 17 cm. Horizontal distance of oblique camera lens from substrate = 122 cm from substrate. With this system only, video images are displayed in real time on a monitor on board the ship. Use of this system was discontinued in favor of GoPro® v3 because management of video cables during deployments was problematic.
3	GoPro® v1 (Great Lakes connecting channels, 2014–2016)	One GoPro® Hero4 camera	Weighted drop line; oblique view	Single Sartek 10-watt dive light	30 frames/second, 1080-pixel resolution, narrow field of view	Height of camera above substrate not controlled. This method was used for the connecting channels pilot studies as part of NCCA 2015 (Pawlowski et al., in prep), and these videos were tested in an early beta test version of DLE.
4	GoPro® v2 (Niagara River)	Two GoPro® Hero5 cameras	Custom video carriage – small; down-looking and oblique view	Single Suptig 84-LED dive light per camera	60 frames/second, 1080-pixel resolution, wide field of view, brightness 80%, otherwise default settings	Lens of down-looking camera 31 cm above substrate. Field of view (video quadrat area) of down-looking camera $\approx 23 \times 54$ cm. Vertical height of oblique camera from substrate = 24 cm. Horizontal distance of lens of oblique camera from substrate = 91 cm.
5	GoPro® v3 (Lake Ontario 2018)	Two GoPro® Hero5 cameras	Custom video carriage large; down-looking and oblique view	Two Suptig 84-LED dive lights per camera	60 frames/second, 1080-pixel resolution	Video carriage is described in Angradi (2018). Lens of down-looking camera was 56 cm above substrate. Field of view (videoquadrat area) of down-looking camera = 42×78

Number	System (application)	Camera system	Deployment method	Lighting	Camera settings	Other details
						cm. Vertical height of oblique camera above substrate = 30 cm. horizontal distance of lens of oblique camera from substrate = 101cm.

Appendix 3. Methods for expert analysis of videos

All the videos were viewed by authors MJW (at least once) and TRA (at least twice), both experienced in identification of *Dreissena* and round goby from videos. Benthic habitat attributes and the presence and relative abundance of invasive species were determined by viewing the full videos on a computer monitor. The two authors independently reviewed Lake Ontario videos. Differences between the two experts' classifications were reconciled by reviewing the video together for a final classification for Lake Ontario, and the Lake Huron and Niagara River videos were classified by TRA with support from MJW. Results for the analysis presented in this paper were based on the expert classification for *Dreissena* and round goby presence, vegetation presence, and substrate type.

The following information was extracted from each video: 1) dominant mineral and biogenic substrates corresponding as closely as possible to the Coastal and Marine Ecological Classification Standard (CMECS; FGDC 2012); 2) the categorical presence (>1 and <5 individuals) or abundance (< 5) of round gobies appearing in the video; 3) the presence (>1 and <10 mussels) or abundance (< 10 mussels) of live *Dreissena* visible in the video; the time (seconds) to detection of the first round goby in video in which they appeared, and the total bottom time (seconds). Our subjective criteria for abundance of *Dreissena* and round gobies was based on the experience of viewing the video many times and making a judgment about what constitutes abundant in the context of all the videos. If we could not make a positive determination of the presence of a round goby or dreissenid mussel in questionable cases, we defaulted to not present.

We did not attempt to identify dreissenid mussels to species from videos. Based on recent publications (Pennuto et al., 2012; Karatayev et al., 2015; Nalepa et al., 2018), we expect that at the depths we collected UVID in Lake Huron and Lake Ontario, the dreissenid fauna was dominated by or exclusively comprised of quagga mussels. In the shallower Niagara River, both quagga and zebra mussels are present (Mehler et al., 2018; Unpublished USEPA data based on PONAR samples collected in 2018).

Because videos provide only visual information, we could not always map CMECS directly to our observations of substrates. Our category of bedrock corresponds to CMECS subclass bedrock. Our category of coarse/gravel approximately corresponds to the CMECS subclass of coarse unconsolidated substrate; our category of fine hard (sand) approximately corresponds to the CMECS substrate groups sand and muddy sand; and our category of fine soft (silt and mud) approximately corresponds to the CMECS substrate group mud. We characterized the dominant mineral substrate in each video according to these classes.

Dreissena reefs formed of self-adhered live mussels correspond to CMECS biogenic substrate group mussel reef. We also noted when mussel shell hash, algae, and submerged leafy aquatic vegetation were present. We then summarized data such that all dominantly bedrock and coarse/gravel were considered “dominantly hard” and all dominantly fine hard and fine soft (sand, silt and mud) were considered “dominantly soft” for comparison to DLE.

We also occasionally observed other organisms in the videos including crayfish, *Mysis*, *Holopedium*, and several species of fish including yellow perch (*Perca flavescens*), alewives (*Alosa pseudoharengus*), rainbow smelt (*Osmerus mordax*), and smallmouth bass (*Micropterus dolomieu*). We did not positively identify any demersal fish species other than round goby. In two videos, the presence of a sculpin (Cottidae) was suspected.

Appendix References:

- Angradi T, 2018 A field observation of rotational feeding by *Neogobius melanostomus*. *Fishes* 3, 5.
- FGDC (Federal Geographic Data Committee). 2012 Coastal and Marine Ecological Classification Standard. 6 2012 https://www.fgdc.gov/standards/projects/cmecs-folder/CMECS_Version_06-2012_FINAL.pdf/view; accessed 01082020
- Great Lakes Sea Grant, 2019 Cooperative Science and Monitoring Initiative (CSMI). <https://greatlakesseagrant.com/projects/cooperative-science-and-monitoring-initiative-csmi/>; accessed 01082020
- Karatayev AY, Burlakova LE, Padilla DK, 2015 Zebra versus quagga mussels: A review of their spread, population dynamics, and ecosystem impacts. *Hydrobiologia* 746, 97–112.
- Kincaid T, Olsen T, Stevens D, Platt C, White D, Remington R, Kincaid MT 2019 Package ‘spsurvey’. <https://cran.r-project.org/web/packages/spsurvey/index.html>; accessed 01082020
- Lietz JE, Kelly JR, Scharold JV, Yurista PM 2015 Can a rapid underwater video approach enhance the benthic assessment capability of the National Coastal Condition Assessment in the Great Lakes? *Journal of Environmental Management* 55, 1446–1456.
- Mehler K, Burlakova LE, Karatayev AY, Biesinger Z, Valle-Levinson A, Castiglione C, Gorsky D, 2018 Sonar technology and underwater imagery analysis can enhance invasive *Dreissena* distribution assessment in large rivers. *Hydrobiologia* 810, 119–131.
- Nalepa TF, Riseng CM, Elgin AK, Lang GA, 2018 Abundance and distribution of benthic macroinvertebrates in the Lake Huron system: Saginaw Bay, 2006–2009, and Lake Huron, including Georgian Bay and North Channel, 2007 and 2012. National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory, Ann Arbor, MI, Technical Memorandum GLERL-172
- Pawlowski M, Wick M, Bolgrien D, Nord M, Launsbach J, Angradi T. In preparation. 2015 National Coastal Condition Assessment (NCCA) and 2014–2016 Connecting Channel Pilot Assessment. A Special Report for the Great Lakes. U.S. Environmental Protection Agency, Duluth, MN, EPA/600/R-19/097
- Pennuto CM, Howell ET, Makarewicz JC, 2012 Relationships among round gobies, *Dreissena* mussels, and benthic algae in the south nearshore of Lake Ontario. *Journal of Great Lakes Research* 38, 154–160.
- Stevens DL Jr., Olsen AR 1999 Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics* 4, 415–428.
- USEPA. 2018 National Coastal Condition Assessment. <https://www.epa.gov/national-aquatic-resource-surveys/ncca>; accessed 01082020

References

- Bell S, Marzano M, Cent J, Kobierska H, Podjed D, Vandzinskaite D, Reinert H, Armaitiene A, Grodzi ska-Jurczak M, Murši R, 2008 What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. *Biodivers. Conserv* 17, 3443–3454.
- Bonney R, Ballard H, Jordan R, McCallie E, Phillips T, Shirk J, Wilderman CC, 2009 Public participation in scientific research: Defining the field and assessing its potential for informal science education. A CAISE Inquiry Group Report. 58 p. <http://www.birds.cornell.edu/citscitoolkit/publications/CAISE-PPSR-report-2009.pdf>, accessed 11/8/2019.
- Bowyer A, Maidel V, Lintott C, Swanson A, Miller G, 2015 This image intentionally left blank: Mundane images increase citizen science participation. In: 2015 Conference on Human Computation & Crowdsourcing. Conference on Human Computation & Crowdsourcing San Diego, California, USA 2 p.
- Brossard D, Lewenstein B, Bonney R, 2005 Scientific knowledge and attitude change: The impact of a citizen science project. *Int. J. Sci. Educ* 27, 1099–1121.
- Burlakova LE, Tulumello BL, Karatayev AY, Krebs RA, Schloesser DW, Paterson WL, Griffith TA, Scott MW, Crail T, Zanatta DT, 2014 Competitive replacement of invasive congeners may relax impact on native species: interactions among zebra, quagga, and native unionid mussels. *PLoS ONE* 9, e114926. [PubMed: 25490103]
- Crall AW, Newman GJ, Stohlgren TJ, Holfelder KA, Graham J, Waller DM, 2011 Assessing citizen science data quality: an invasive species case study. *Conserv. Lett* 4, 433–442.
- Cox J, Oh EY, Simmons B, Lintott C, Masters K, Greenhill A, Graham G, Holmes K, 2015 Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Comput. Sci. Eng* 17, 28–41.
- Di Salvo R, Giordano D, Kavasidis I, 2013 A crowdsourcing approach to support video annotation. VIGTA '13 Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications, 8. St. Petersburg, Russia 6 p. <http://groups.inf.ed.ac.uk/f4k/PAPERS/VIGTA13.pdf>, accessed 11/8/2019.
- Dickinson JL, Shirk J, Bonter D, Bonney R, Crain RL, Martin J, Phillips T, Purcell K, 2012 The current state of citizen science as a tool for ecological research and public engagement. *Front Ecol. Environ* 10, 291–297.
- Environment and Climate Change Canada (ECCC) and the U.S. Environmental Protection Agency (USEPA)., 2018 Lake Huron Lakewide Action and Management Plan, 2017–2021. Cat. No. En164–56/2018E-PDF, ISBN 978–0–660–25841–6. <https://www.epa.gov/sites/production/files/2018-04/documents/lake-huron-lamp-2017-2021-105pp.pdf>, accessed 11/8/2019.
- Eveleigh A, Jennett C, Blandford A, Brohan P, Cox AL, 2014 Designing for dabblers and deterring drop-outs in citizen science. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2985–2994. ACM Toronto, Ontario, Canada 2985 – 2994.
- Developers FFmpeg, 2016 ffmpeg tool (Version be1d324) [Software]. Available from <http://ffmpeg.org/>
- Ferrini VL, Pawlenko N, Morton JJ 2018 Engaging citizen scientists in ocean exploration with Ocean Video Lab. AGU Fall Meeting Abstracts.
- Fitzpatrick MC, Preisser EL, Ellison AM, Elkinton JS, 2009 Observer bias and the detection of low-density populations. *Ecol. Appl* 19, 1673–1679. [PubMed: 19831062]
- Garcia-Molina H, Joglekar M, Marcus A, Parameswaran A, Verroios V, 2016 Challenges in data crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering*. 28, 901–911.
- Gardiner MM, Allee LL, Brown PM, Losey JE, Roy HE, Smyth RR, 2012 Lessons from lady beetles: Accuracy of monitoring data from US and UK citizen-science programs. *Front. Ecol. Environ* 10, 471–476.
- Good BM and Su AI, 2013 Crowdsourcing for bioinformatics. *Bioinformatics*. 29, 1925–1933. [PubMed: 23782614]
- Happel A, Murchie KJ, Willink PW, Knapp CR, 2020 Great Lakes Fish Finder App; a tool for biologists, managers, and education practitioners. *J. Great Lakes Res* 46, 230–236.

- Jiménez M, 2018 A first approach for handling uncertainty in citizen science. In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) IEEE Rio de Janeiro, Brazil 1–8.
- Johnson TB, Bunnell DB, Knight CT, 2005 A potential new energy pathway in Central Lake Erie: the round goby connection. *J. Great Lakes Res* 31, 238–251.
- Kamar E, Hacker S, Horvitz E, 2012 Combining human and machine intelligence in large-scale crowdsourcing In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems. 1, 467–474. International Foundation for Autonomous Agents and Multiagent Systems.
- Karatayev AY, Burlakova LE, Padilla DK, 2015 Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts. *Hydrobiologia* 746, 97–112.
- Karatayev AY, Mehler K, Burlakova LE, Hinchey EK, Warren GJ, 2018 Benthic video image analysis facilitates monitoring of *Dreissena* populations across spatial scales. *J. Great Lakes Res* 4, 629–638.
- Kelling S, Johnston A, Hochachka WM, Iliff M, Fink D, Gerbracht J, Lagoze C, La Sorte FA, Moore T, Wiggins A, Wong WK, 2015 Can observation skills of citizen scientists be estimated using species accumulation curves? *PLoS ONE* 10, e0139600. [PubMed: 26451728]
- Kendall S 1974 Preference for intermittent reinforcement. *J. Exp. Anal. Behav* 21, 463. [PubMed: 16811758]
- Klein A, Silvester S, Tanbakuchi A, Müller P, Nunez-Iglesias J, Harfouche M, McCormick M, Ladegaard A, Rai A, OrganicIrradiation, Smith TD, Konowalczyk M, rreilink, Nises J, jackwalker64, Vaillant GA, Barnes C, Zulko, Hsieh P-C, NiklasRosenstein, Hirsch M, Schambach M, Singleton J, van Kemenade H, Inggs G, Yang G, Kohlgrüber F, Dusold C, Levsakaya A, Elliott A 2019 imageio/imageio: V2.6.0 (Version v2.6.0). Zenodo 10.5281/zenodo.3475011
- Kornis MS, Mercado-Silva N Vander Zanden MJ, 2012 Twenty years of invasion: a review of round goby *Neogobius melanostomus* biology, spread and ecological implications. *J. Fish Biol* 80, 235–285. [PubMed: 22268429]
- Lederer AM, Janssen J, Reed T, Wolf A, 2008 Impacts of the introduced round goby (*Apollonia melanostoma*) on dreissenids (*Dreissena polymorpha* and *Dreissena bugensis*) and on macroinvertebrate community between 2003 and 2006 in the littoral zone of Green Bay, Lake Michigan. *J. Great Lakes Res* 34, 690–697.
- Lietz JE, Kelly JR, Scharold JV, Yurista PM, 2015 Can a rapid underwater video approach enhance the benthic assessment capability of the national coastal condition assessment in the Great Lakes? *Environ Manage.* 55, 1446–1456. [PubMed: 25832344]
- Locke CM, Anhalt-Depies CM, Frett S, Stenglein JL, Cameron S, Malleshappa V, Peltier T, Zuckerberg B, Townsend PA, 2019 Managing a large citizen science project to monitor wildlife. *Wildlife Soc B.* 43, 4–10.
- Madenjian CP, Stapanian MA, Witzel LD, Einhouse DW, Pothoven SA, Whitford HL, 2011 Evidence for predatory control of the invasive round goby. *Biol. Invasions* 13, 987–1002.
- Masters K, Oh EY, Cox J, Simmons B, Lintott C, Graham G, Greenhill A Holmes K, 2016 Science learning via participation in online citizen science. *J. Sci. Comm. Special Issue: Citizen Science, Part II.* arXiv:1601.05973. 32 p.
- Matabos M, Hoeberechts M, Doya C, Aguzzi J, Nephin J, Reimchen TE, Leaver S, Marx RM, Albu AB, Fier R, Fernandez-Arcaya U, 2017 Expert, crowd, students or algorithm: Who holds the key to deep-sea imagery 'big data' processing? *Methods Ecol. Evol* 8, 996–1004.
- Mayer CM, Burlakova LE, Eklöv P, Fitzgerald D, Karatayev AY, Ludsin SA, Millard S, Mills EL, Ostapenya AP, Rudstam LG, Zhu B, 2014 Benthification of freshwater lakes: exotic mussels turning ecosystems upside down, in: *Quagga and Zebra mussels: Biology, Impacts, and Control*, 2nd ed Lewis Publishers, Boca Raton, FL, 575–586.
- Mehler K, Burlakova LE, Karatayev AY, Biesinger Z, Valle-Levinson A, Castiglione C, Gorsky D, 2018 Sonar technology and underwater imagery analysis can enhance invasive *Dreissena* distribution assessment in large rivers. *Hydrobiologia* 810, 119–131.
- Mugar G, Østerlund C, Hassman KD, Crowston K, Jackson CB, 2014 Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. In:

Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing ACM Baltimore, Maryland, USA 109–119.

- Nov O, Arazy O, Anderson D, 2014 Scientists@ Home: what drives the quantity and quality of online citizen science participation? PloS ONE 9, e90375. [PubMed: 24690612]
- Ozersky T, Barton DR, Evans DO, 2011 Fourteen years of dreissenid presence in the rocky littoral zone of a large lake: Effects on macroinvertebrate abundance and diversity. J. N. Am. Benthol. Soc 30, 913–922.
- Pawlowski M, Wick M, Bolgrien D, Nord M, Launspach J, Angradi T, in preparation. 2015 National Coastal Condition Assessment (NCCA) and 2014–2016 Connecting Channel Pilot Assessment A Special Report for the Great Lakes. U.S. Environmental Protection Agency, Duluth, MN, EPA/600/R-19/097. 77 p.
- R Core Team, 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org/>
- Raddick MJ, Bracey G, Gay PL, Lintott CJ, Murray P, Schawinski K, Szalay AS, Vandenberg J, 2009 Galaxy zoo: Exploring the motivations of citizen science volunteers. Astr. Educ Rev 9, 010103 arXiv:0909.2925.
- Rahman M, Blackwell B, Banerjee N, Saraswat D, 2015 Smartphone-based hierarchical crowdsourcing for weed identification. Comput. Electron. Agr 113, 14–23.
- Reback J, McKinney W, Van den Bossche J, jbrockmendel, Augspurger T, Cloud P, gyoung, Sinhrks, Klein A, Tratner J, She C, Roeschke M, Ayd W, Petersen T, Hayden A, Hawkins S, Schendel J, Garcia M, Jancauskas V, Battiston P, Seabold S, chris-b1, h-vetinari, Hoyer S, Overmeire W, Mehyar M, Whelan C, behzad nouri, Kluyver T, Chen KW 2019 pandas-dev/pandas: v0.25.3 (Version v0.25.3). Zenodo 10.5281/zenodo.3524604
- Rotman D, Preece J, Hammock J, Procita K, Hansen D, Parr C, Lewis D, Jacobs D, 2012 Dynamic changes in motivation in collaborative citizen-science projects. In: Proceedings of the ACM 2012 conference on computer supported cooperative work ACM Seattle, Washington, USA 217–226.
- Salk CF, Sturm T, See L, Fritz S, 2017 Limitations of majority agreement in crowdsourced image interpretation. T. GIS. 21, 207–223.
- Salman A, Jalal A, Shafait F, Mian A, Shortis M, Seager J, Harvey E, 2016 Fish species classification in unconstrained underwater environments based on deep learning. Limnol. Oceanogr-Meth. 14, 570–585.
- Stanley J 2017 tidyjson: A Grammar for Turning ‘JSON’ into Tidy Tables. R package version 0.2.2 <https://www.rdocumentation.org/packages/tidyjson/versions/0.2.2>
- Swanson A, Kosmala M, Lintott C, Packer C, 2016 A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conserv. Biol 30, 520–531. [PubMed: 27111678]
- Torre M, Nakayama S, Tolbert TJ, Porfiri M, 2019 Producing knowledge by admitting ignorance: Enhancing data quality through an “I don’t know” option in citizen science. PloS ONE 14, e0211907. [PubMed: 30811452]
- Trebitz AS, Hatzenbuehler CL, Hoffman JC, Meredith CS, Peterson GS, Pilgrim EM, Barge JT, Cotter AM, Wick MJ, 2019 Dreissena veligers in western Lake Superior–Inference from new low-density detection. J. Great Lakes Res 45, 691–699. [PubMed: 31359907]
- Zooniverse, 2019 Welcome to the Zooniverse. <https://www.zooniverse.org>, accessed 11/8/2019.
- Zulko, 2017 MoviePy (version 1.0.1). <https://pypi.org/project/moviepy/>

Figure 1.
Map of sites in Lake Huron, Lake Ontario, and the Niagara River where UVID was collected. In Lake Huron, 71 videos were collected within 2 km of the 8 base sites shown.

Figure 2.

Video carriages used for collection of UVID. The larger carriage, left, measuring $61 \times 61 \times 91$ cm, was deployed in the Great Lakes from a ship, and the smaller carriage, right, measuring approximately $60 \times 60 \times 60$ cm, was deployed in the Niagara River from a boat. Down-looking and oblique video were collected, but only oblique video was classified in Deep Lake Explorer.

Figure 3.

Percent DLE agreement with expert classification for videos with a range of volunteer agreement thresholds applied to determine DLE video result. Vertical lines represent the thresholds applied for reporting results. For example, applying a threshold of 40% volunteer agreement that vegetation was present, 90% of videos classified by volunteers in DLE had the same result as the expert. We selected thresholds by the (1) highest total percent agreement with expert results (solid line), (2) lowest false positives (present based on DLE, absent based on expert classification; dashed line), and (3) lowest false negatives (absent based on DLE, present based on expert classification; dot-dot-dashed line). DLE-expert classification agreement at the selected thresholds are given in Fig. 4.

Figure 4.

Percent volunteer agreement for true negatives, true positives, false negatives, false positives, all correct, and all incorrect for four attributes. For *Dreissena*, vegetation presence, and substrate (shown as hard substrate), the vertical axis is mean percent volunteer agreement for the videos in each group ($\pm 95\%$ CI). For round gobies, the vertical axis is the mean of the maximum percent volunteer agreement for the clips of the videos in each group. The number above each bar is the percent of the total dataset represented by each bar.

Table 1.

Number of clips and videos classified in DLE, number compared to expert classification, and sites at which UVID was collected. Some sites had multiple drops, so there were more videos collected than sites. Slightly fewer videos were compared to expert than classified in DLE because 3 poor quality videos included in DLE were not classified by the expert and were therefore not included in the DLE-expert comparison.

Waterbody	Clips classified in DLE	Videos classified in DLE	Videos compared to expert	Sites collected
Lake Huron	371	61	59	59
Niagara River	196	86	86	69
Lake Ontario	179	52	51	34
Total	746	199	196	162

Table 2.

List of questions about attributes in UVID clips in the order posed in the DLE workflow. Volunteer agreement threshold refers to the percent of volunteers that must agree on the positive identification (or left-most result in this table) for that result to be considered true. Vegetation was defined as any vegetation taller than a surface-coating fuzz (e.g., periphyton). Hard bottom was defined as dominantly coarse material including rock, boulders, gravel, and/or cobble; soft bottom was defined as fine material including sand, silt, and/or clay. Appendix 1 video 7 shows a walk-through of DLE.

	Question	Possible responses		Volunteer agreement threshold
1.	Do you see any fish?	Yes	No	30%
2.	Are any of the fish round goby?	Yes	No	30%
3.	Do you see any vegetation?	Yes	No	40%
4.	Do you see any zebra or quagga mussels?	Yes	No	40%
5.	Are mussels abundant and covering more than half the area visible in the video?	Yes	No	40%
6.	What is the dominant substrate?	Hard bottom	Soft bottom	50%
7.	Do you see any human-made features, objects, or items?	Yes	No	40%
8.	Was the quality/resolution of this video adequate to complete the classifications?	Yes	No	30%

Table 3.

Agreement between two experts' classifications, accuracy of DLE classification, accuracy of DLE classification with clips removed that had 70% volunteer agreement that the clip was poor quality (29 of 746 clips), and detection rates for each attribute classified by both DLE and experts. Detection rate is the number of videos with attribute detected by both DLE and expert/number of videos where expert detected the attribute. Detection rates are given for hard substrate/soft substrate.

	Agreement (%) between two experts (Lake Ontario only, N = 51)	Agreement (%) between DLE and expert N = 196	Agreement (%) between DLE and expert, removing the 29 poorest-quality clips N = 196	Detection rate (%) N = 196
Round goby presence	96	89	90	78
<i>Dreissena</i> presence	92	79	80	66
Vegetation presence	94	90	90	93
Dominant substrate	90	77	77	62/95

Table 4.

Results of videos classification from Deep Lake Explorer and by an expert. Values are percent of videos. Substrate type is given as percent dominantly hard; the remaining proportion of videos assessed were classified as dominantly soft.

	Lake Huron		Niagara River		Lake Ontario		All	
	DLE	Expert	DLE	Expert	DLE	Expert	DLE	Expert
Round gobies present	70	76	5	15	46	61	36	46
<i>Dreissena</i> present	31	55	23	42	71	84	38	57
Vegetation present	34	27	80	74	8	8	47	43
Hard substrate	44	58	37	69	15	22	34	53