AI proof-of-concept for counting mussels

Context:

Since their accidental introduction in the late 1980s, the *dreissenid* zebra and quagga mussels have colonized large portions of the bottoms of most of the Great Lakes. This invasion does not only cause a physical nuisance, but it also has bioengineered the lakes to the extent that ecosystems have been disrupted and established nutrient pathways have been irreparably altered. Mapping the presence and density of zebra mussels using temporal density maps is vital to monitoring and determining the impacts of this species invasion. Existing maps are based on limited site measurements, which require costly sampling and laboratory analysis. Currently, there is a lack of maps showing the full spatial extent of these invasive mussels in Canadian lakes such as the Great Lakes and Lake Winnipeg. It is expected that AI analyses of underwater video can be used to effectively and efficiently enumerate the density and distribution of mussels on the lake bottom. This proof-of-concept will be on the feasibility of combining AI with underwater imaging to count mussels and estimate their density and biomass.

Data:

Still images were collected by divers between 2012 and 2018 when mussel samples were taken at several sites in Lake Ontario and Lake Erie (total: 1,608 analyses) with analyses on percentage coverage, density, biomass, and viability. The mussel analyses at these sites can be used to train and validate the performance of the AI algorithm. New stills, extracted from underwater video, are planned to be acquired in late October 2019 to support the ongoing research in the eastern Lake Erie. The 2019 images will be recorded at a relatively constant altitude above the lake bottom and include laser spacing to record length scale in each image. The determination of length scale can be applied to size the mussels and estimate their coverage area.

Scope-of-work:

- Pre-process underwater mussels image data for ingestion by computer vision model:
    a) Cropping of image area inside "quadrat."
    b) Any other pre-processing steps necessary before ingestion by computer vision model
- Train the computer vision model on underwater mussels image for:
    a) Percentage live coverage
    b) Total mussels count
    c) Estimated biomass
- Validate the computer vision model according to relevant metrics:
    a) Leave-one-out-cross-validation
    b) Cross-validation between different years and between different sites
- Perform prediction on new mussels image data acquired in 2019
- Prepare and edit as requested a PowerPoint presentation and a written report
- Present work at CCIW at mid-contract and end-of-contract stage
- Be available for a bi-weekly 1-hour phone conversion to report on progress as needed

Deliverables:

- Commented and modularized Python code using Open Source packages for pre-processing, training, prediction and validation of the computer vision model
- CSV file containing predicted percentage coverage, mussels count and estimated biomass for each image for stills acquired from divers and for new images acquired from a camera system deployed from a boat
- CSV file containing error estimates for computer vision model for both training and validation steps
- Written report documenting the steps taken, the rationale for the various design choices made and parameters selected for the training of the computer vision model, summarizing the strengths and limitations of the computer vision model, and detailing how data (video) acquisition methods can be improved for future surveys.
- PowerPoint presentation that can be used for presentation to management. It should provide an overview of the process, the selected method and outcomes and include examples of images and results.

Milestones:

- November 29th, 2019: Pre-processing step, validation infrastructure ready with a trivial computer vision model
- December 20th, 2019: Training and validation of the computer vision model for percentage live coverage
- January 10th, 2020: Training and validation of the computer vision model for mussels count and biomass
- January 17th, 2020: Prediction using the 2019 underwater images
- January 31st, 2020: Submission of the report and final presentation

Innovate with AI at ECCC Dragon's Den

Proposal Submission Form

1.	**Name of person submitting use case**
	Dominique Brunet, Reza Valipour, Johann (Hans) Biberhofer
2.	**Name of use case sponsor (Director General or higher)**
	Dr. Kevin Cash
3.	**Name of team/branch**
	Team mussels/Water S&T
4.	**Please provide your contact information (email or phone number) should we have questions or require more information about your use case.**
	Dominique.Brunet@canada.ca
5.	**Please clearly describe the use case/business problem you are trying to solve, why an AI solution would be the most effective approach to address it and the impact the AI solution would have on your work. We encourage you to use the Treasury Board of Canada Secretariat's Algorithmic Impact Assessment to help determine the potential impact of AI on your use case (https://open.canada.ca/aia-eia-js/?lang=en).**
	Since their accidental introduction in the late 1980s, the dreissenid zebra and quagga mussels have colonized large portions of the bottoms of most of the Great Lakes. This invasion does not only cause a physical nuisance, but it has also bioengineered the lakes to the extent that ecosystems have been disrupted and established nutrient pathways have been irreparably altered. Mapping the presence and density of zebra mussels using temporal density maps is vital to monitoring and determining the impacts of this species invasion. Existing maps are based on limited site measurements which requires costly sampling and laboratory analysis. Currently there is a lack of maps showing the full spatial extent of these invasive mussels in Canadian lakes such as the Great Lakes and Lake Winnipeg. Underwater images or videos of mussels have been collected with high definition cameras deployed from ships. Potential also exists to use camera systems integrated into Remotely Operated underwater Vehicles (ROVs) or Autonomous Underwater Vehicles (AUVs). While existing video assets provide useful qualitative information, quantitative mussel density analysis requires time-consuming, costly and often error-prone manual counting of mussels. For example, three weeks of field sampling can provide detailed information for 36 sites, costs about 17,000$ in boat and travel time, requires 4 to 5 persons in the field to collect the mussels, and 7 persons in the lab to receive, freeze-dry, clean, run through the sieve for size sorting, count, and weigh the mussels. The AI solution would provide a mean to quickly estimate count, size and weigh of mussels from images or videos without the need for lab analysis. Depending on the available imaging platform, it could also be possible to reduce the need for divers at every site and to extend the areal coverage of the mussel mapping. Beyond potential for cost and time savings, the impact of the AI solution would be to enable new capabilities that are essential for downstream applications. Most importantly, the delineation of mussel densities is an essential input for lake-wide or regional numerical water quality models to study the impact of mussels on nutrient recycling. A comparison of the mussel density between different years will provide relative changes of mussel populations which in turn will support integrated science in assessing the role these invasive species play as ecosystem engineers and modifiers of nutrient dynamics that fuel harmful algal blooms. Finally, based on lessons learned similar AI algorithms could be developed for other purposes such as substrate type classification, identification of green algae, and counting of invasive fish species. This proposal will thus advance the scientific capacity to define status and assess management of nutrients for water bodies impacted by mussels in a cost-efficient way.

6.	**If you completed TBS' Algorithmic Impact Assessment, please provide the score you received for your proposed use case.**
	Impact Level 2; Current Score 29; Raw Impact Score 29; Mitigation Score 12
7.	**Please describe how you have consulted with appropriate groups (CIO, CDO, subject matter experts, policy and program leads, data stewards, etc.), how they informed your proposed use case and how you addressed their input and advice.**
	We have initiated discussions with program leads (Dr. Ram Yerubandi and Dr. Mohamed Mohamed). We have also contacted Dr. David Depew, subject matter expert in aquatic biology, who suggested to look at a paper in which a survey of percentage coverage of dreissenid mussels in Lake Simcoe, Ontario was performed from manual analysis of videos at several sampling sites. We then discussed with Chris Duggan, lead diver, who described manual sampling and imaging methods for current sites as well as with Leah Peacock, scientific technologist, and Megan McCusker, physical science specialist, who informed us on the current (manual) sample analysis and data preparation methods. Megan McCusker also helped with the cost estimate and suggested to look at work of other researchers in US. Finally, Jacqui Milne from Water S&T provided suggestions on how to publish added-value data on the Open Data Portal. We noted all inputs and we will seek to initiate collaborations once the proof of concept is approved. We also consulted with the CDO and CIO and we followed suggestions to reduce the amount of details in the proposal and to put a greater emphasis on value-for-money impact.
8.	**Please describe your willingness to carry out the proof of concept and implement the AI solution should the proof of concept be successful (specifically identifying the resources available to support the completion of the proof of concept, as well as scaling it if it is successful).**
	We are very motivated to implement an AI solution for the automated analysis of images and videos. The proof of concept will be mainly supported by Dr. Dominique Brunet, with field data support from Johann Biberhofer for underwater imaging and oversight by Dr. Reza Valipour to ensure that the products contribute value-added to existing water quality models. Based on initial trials we are confident that the proof of concept will be successful. The proof of concept will initially apply the AI solution to images and videos encompassing the northern reaches of east Lake Erie. The scaling up will involve collaboration with other researchers in Environment and Climate Change Canada and other departments or agencies on both sides of the Great Lakes who are interested or engaged in determining the distribution and densities of dreissenid mussel populations. The lessons learned here will be transferrable to other Canadian lakes where there are existing or emerging issues related to dreissenid mussels such as in Lake Winnipeg. Moreover, in the future the AI solution could be further developed to include substrate types (sand, mud, gravel, rock, etc.), *Cladophora* (green algae) or the count of round goby fish.
9.	**Please describe whether there is enough usable, discoverable and available data to use for the proof of concept on this use case.**
	A unique opportunity exists to compare spatially referenced underwater video of bottom substrate, including dreissenid mussels, collected in 2006 with new videos planned to be acquired in October 2019 to support the ongoing research in the eastern Lake Erie. The 2006 videos were recorded using standard definition underwater cameras deployed from surface vessels. As a consequence, the height above the bottom was variable. The 2019 videos will be recorded at a relatively constant depth and include laser spacing to record length scale in each frame. The determination of length scale will be used to size the mussels and estimate their coverage area. There are additional videos and still images between 2012 and 2018 when mussel samples were taken at several sites in Lake Ontario and Lake Erie (total: 1,608 analyses) with analyses on percentage coverage, density, biomass, and viability. The mussel analyses at these sites will be used to train and validate the performance of the AI algorithm.