Software Competition Helps Researchers Tell Right Whales Apart – by their Head Lice

There are only about 500 right whales in the North Atlantic, and telling these endangered animals apart using photographs taken from an aircraft flying 750 feet overhead is a huge challenge. Even with the help of a catalog maintained by the New England Aquarium, it takes many hours to match the images taken on each flight to those in the catalog.

Christin Khan, a fishery biologist in the Northeast Fisheries Science Center (NEFSC)'s Protected Species Branch, and colleagues in the NEFSC's aerial survey group spend hundreds of hours each year flying in the NOAA Twin Otter on surveys off the Northeast coast, looking primarily for right whales. Depending on how many whales are sighted, dozens to hundreds of photographs can be taken on a typical flight lasting six hours.

How do they do it? North Atlantic right whales have distinctive patterns of whale lice known as callosities on their heads. Those patterns, much like human fingerprints, help identify one individual right whale from another. Khan wondered how she could develop software that would recognize the callosity patterns and free up time so she and her colleagues could focus on other right whale research needs and priorities.

"We have a very real problem and needed a solution," Khan said of the challenge, which she first began thinking about how to solve in the fall of 2013.

The numbers of aerial survey images needing identification is daunting, and that assumes right whales are in the area. In 2011, for example, 4,300 frames contained 349 individuals and led to 329 identifications. In 2015, when few right whales were observed during aerial surveys, 1,360 frames revealed 102 individuals, of which 83 were identified. While some images can be identified quickly, many cannot and the task can be tedious. Images may contain several animals, the angle and clarity of the image might not be suitable for an identification, and further work might be required to determine a match or to identify a whale that is not in the catalog.

Khan initially wanted an algorithm that would help identify right whales from images collected from any platform, whether an aircraft or a vessel on the surface, since she and colleagues are involved in both types of surveys. After learning the computer processes were very different for each perspective, she decided to focus on images taken from aircraft.

After pursuing several leads in early 2014, Khan found Kaggle, an online data analytics competition site. The site had been recommended by Cornell University scientist Christopher Clark, who studies whale sounds and was involved in a right whale call detection algorithm contest himself.

Kaggle agreed to launch a competition for an algorithm to identify right whales, working with Khan to frame the parameters of the competition and using 4,500 labeled aerial photographs of individual whales supplied by Khan and colleague Leah Crowe as the data set. MathWorks, a technical computing software company based in Massachusetts and known for its MATLAB software, offered $10,000 in prize money and free software to all participants.

The contest began August 15, 2015 and attracted 364 teams with 470 players from around the world before it ended January 7, 2016. During the challenge, Khan checked in on the Kaggle forums to provide basic information or answer questions from competitors curious about what the whales were doing in some photos or about basic whale anatomy.

The winning team, from the Warsaw, Poland office of the international data science company deepsense.io, developed an algorithm that can identify whale "faces" with 87 percent accuracy. They received $5,000.

"This was an amazing problem to tackle and an amazing competition to take part in," Robert Bogucki and other deepsense.io team members noted in a blog post explaining how they arrived at their solution. "We have learned a lot during the process, and are actually quite amazed with the superpowers of deep learning!"

The second place winner, Felix Lau, was an individual contestant. Lau was attracted to this challenge because it was a unique opportunity and he "could help save a species. I wanted to apply the latest deep learning techniques and ideas I had learned in the past few months to a real problem, and the right whale recognition challenge came at the right time." He received $3,000.

When not involved in competitions like Kaggle, Lau said he tries to read a few papers a week and catch up with the latest advancements in the field. "Winning one of the top prizes gave me confidence that not everyone needs a Ph.D. or have a research background (like myself) to do well in the field of machine learning. It was particularly encouraging to know that my effort to learn about the latest computer vision developments did not go to waste."

Lau had participated in two previous Kaggle image competitions, one on diabetic retinopathy detection and the other the first national data science bowl. While he didn't do as well in those efforts, the experiences gave him ideas on how to approach these kinds of competitions and allowed him to get started on the right whale recognition challenge quickly.

The third place winner of $2,000 is SKE, a trio of researchers from the U.S. and Japan whose team name represents the first letter of their first names. The team includes Shize Su, a Ph.D. student in the electrical and computer engineering department at the University of Virginia; Kohei Ozaki, a data scientist in Tokyo, Japan; and Eben Olson, an associate research scientist in biomedical imaging at Yale University.

Shize had learned a lot from many past Kaggle competitions and wanted to learn more about image processing from his teammates through their collaboration, while Kohei said he entered the competition "just for fun." Eben Olson, who joined the team late in the process, has competed in a number of Kaggle competitions. Each team member spent between 1-4 hours a day on the competition, and indicated they learned from each other.

"The cool part of the project was the way we were able to solve the problem," Khan said of the Kaggle competition. "Some of the best data scientists in the world came together to tackle a problem that is important in right whale conservation. The participants freely shared their ideas and frustrations along the way and helped each other. I think they were intrigued by the challenge, and it was a way for them to learn something new and push the boundaries."

Khan hopes the open source algorithm will be put to use within a year, by the start of the next calving season. At the moment, she is looking forward to working with the algorithm to see if the NEFSC can write the software using the algorithm in house, or partner with external collaborators to incorporate it into open source software widely available to the research and conservation community and others interested in protecting right whales.

Further down the road, she plans to talk with colleagues in the right whale research community about whether they should seek another algorithm to identify whales from photos taken from vessels and combine the two into a single software application.

And while the Kaggle campaign has been a rewarding experience, it has not been the only effort Khan has pursued in the past year or so to help right whales. Beside aerial survey flights and other research duties, she spends time making the public aware of the presence of right whales through a sign campaign.

Ship strikes are the primary cause of death for right whales, leading Khan and collaborators to make and place signs in marinas along the East coast, informing boaters going offshore about what to do if they see a right whale and how they can help save them. Since September 2014, more than 75 signs have been posted at boat ramps and marinas from Maine to Virginia, with more planned.

As for the Kaggle competition algorithm, Khan says it will be a big help to right whale researchers, who will be able to increase their efficiency and productivity with the time-saving software. Some other possible applications are in genetic studies, sound recordings, and disentanglement efforts where identification of the individual animal involved is important.

"The competition didn't cost us anything, other than some time to pull together the data set," Khan said. "It was an amazing experience, and I would encourage others to take advantage of this innovative and efficient way to solve a problem. There is a real opportunity here to accomplish great things with limited resources. We are excited to get the algorithm turned into software that scientists can take into the field. "

# # #

Related links:

NEFSC Whale Research: http://www.nefsc.noaa.gov/psb/whales/

Kaggle competition: https://www.kaggle.com/c/noaa-right-whale-recognition/details/prizes

Right Whale Callosity pattern identification:
http://www.neaq.org/conservation_and_research/projects/endangered_species_habitats/right_whale_research/right_whale_projects/monitoring_individuals_and_family_trees/identifying_with_photographs/how_it_works/callosity_patterns.php

North Atlantic Right Whale Catalog: http://rwcatalog.neaq.org/Terms.aspx

Interactive Map of North Atlantic Right Whale Sightings: http://www.nefsc.noaa.gov/psb/surveys/

Competition Winner's solution: http://deepsense.io/deep-learning-right-whale-recognition-kaggle/

Second place winner's solution (blogpost): http://felixlaumon.github.io/2015/01/08/kaggle-right-whale.html

Making Facebook for Whales: http://www.theatlantic.com/science/archive/2016/01/north-atlantic-right-whales-facial-recognition/424113/

New Whale Species, and New Ways to Study Them All (article and audio):
http://capeandislands.org/post/new-whale-species-and-new-ways-study-them-all

Collaboration Key to New Right Whale Sign Campaign:
http://www.nefsc.noaa.gov/news/features/rw_sign/