

THE AUTHOR FILE

Juan C. Caicedo

A machine-learning competition in microscopy and the fun of gazing into the night sky.

“I really like algorithms and interesting solutions, hard problems to crack,” says computer scientist Juan Caicedo, who is starting his lab at the Broad Institute of MIT and Harvard. He is a Schmidt Fellow, a program that appoints researchers in physics and computer science to become principal investigators and take on tough problems in biology. One such tough problem is identifying cellular structures in micrographs.

Before becoming a Schmidt Fellow this summer, Caicedo completed his PhD at Universidad Nacional de Colombia, where the community around him supported research, conferences and internships in academic and industrial labs. As a result, his record enabled postdoctoral positions at the University of Illinois at Urbana–Champaign, and in Anne Carpenter’s lab at the Broad.

As a graduate student, Caicedo had helped pathologists at an imaging center extract information from stained tissue slides. The algorithms were not quite up to the task, he says. “At that time, I thought there is a lot of room for improvement.” Image analysis in biology had him hooked.

Now it’s 2019, and computational face-detection algorithms are common. But in cell biology, many labs manually segment nuclei in their micrographs. That can seem safer and easier than configuring software that remains not quite applicable to the work, he says. “Maybe one of the things that precludes people from using imaging for their biological experiments is the question of ‘OK, who is going to analyze the images after we acquired them?’” He and his colleagues now present results of the 2018 Data Science Bowl, a competition of microscopy segmentation algorithms.

Around 18,000 participants, in nearly 3,900 teams, took on the challenge of detecting stained nuclei entirely without human intervention in many types of 2D microscopy images. One of the competition stages asked participants to segment around 100,000 nuclei in 3,200 images in seven days. “We were generally surprised by the participation that this competition attracted,” he says. There were prizes for the top five teams, including cash and an NVIDIA workstation. “Deep convolutional networks were the best solution in this competition,” says Caicedo. They’re a class



Juan C. Caicedo.

of algorithms used in computer vision and natural language processing. The teams trained their algorithms on datasets that Caicedo and colleagues had compiled of nearly 38,000 manually annotated nuclei in over 800 images, from more than 30 experiments across different cell lines, imaging conditions, research facilities and protocols.

The idea for the competition arose in a brainstorming jamboree in the Carpenter lab in 2017. The team decided to run a competition on Kaggle, the Google-owned data science site where teams compete to, for example, classify forests from land-cover data or identify humpback whales from fluke photos. The Data Science Bowl winners brought a variety of approaches to the challenge, which sets the stage for the community’s next steps. In parallel, Caicedo and his colleagues also tested U-Net, a commonly used algorithm, for its efficiency in segmenting nuclei in micrographs. “It’s designed to work with little data, but at the same time it doesn’t have the capacity to work with a variety of data,” he says.

All of the teams in the competition have made their tools publicly available but “there’s still a way to go in order to get something that works for everybody,” he says; and the tools cannot yet take on the diversity of images and tasks in cell biology labs. Microscopy image analysis remains the works of experts and so much remains to be discovered about cells, he says. Rather than a conclusion, the Data Science Bowl results are a proof of concept and an indication of work to be done.

Among other data analysis challenges in biology, Caicedo plans to continue working on approaches in automated microscopy image analysis. “We need more images,” he says, to improve the training sets for algorithms. Manual annotation of these data is challenging, as is finding images for scale-up, given that labs might not readily wish to share them. But he will keep going. “One of the next frontiers: can we add more cell structures into the data set?” he says. Such algorithms could recognize nuclei, cytoplasm or organelles.

Caicedo brings determination to his pastimes, too. “I bike to work every day and I try to do it no matter the weather,” he says. He enjoys traveling, especially if it involves celestial events. “I really enjoy astronomy,” he says. He likes visiting astronomy centers and peering at the night sky through telescopes.

“I really like algorithms and interesting solutions, hard problems to crack.”

Lassi Paavolainen, a postdoctoral fellow at the Institute for Molecular Medicine Finland admires Caicedo’s constant positivity in life and enthusiasm for science. “He is a truly great scientist but also a really nice guy to talk to and work with,” says Paavolainen, who collaborates with Caicedo. The two researchers “clicked” over a soccer conversation several years ago and deepened their friendship during Paavolainen’s three-month stint at the Broad last fall. “Juan has been one of the key forces in recent developments in image-based profiling, mainly in utilization of state-of-the-art deep-learning approaches for segmentation and representation learning,” he says.

Paavolainen looks forward to their next collaborations and his colleague’s trajectory. The Broad “definitely made the right decision granting him a Schmidt fellowship.” □

Vivien Marx

Published online: 19 November 2019
<https://doi.org/10.1038/s41592-019-0668-4>

Reference

Caicedo, J. C. et al. Nucleus segmentation across imaging experiments. *Nat. Methods*. <https://doi.org/10.1038/s41592-019-0612-7> (2019).