

CELL BIOLOGY

Raising the game in image classification

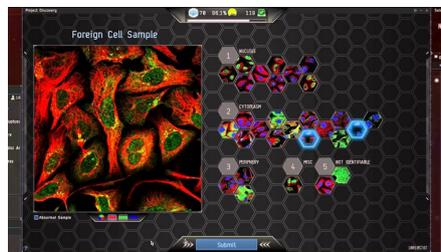
By incorporating an image-classification task into an online video game, the Human Protein Atlas project unlocks the gaming community as a citizen science force.

The Cell Atlas project aims to generate a map of the human proteome at subcellular resolution. This project is part of a larger effort, the Human Protein Atlas (HPA), which includes the Tissue Atlas and the Pathology Atlas. “The long-term goal of the Cell Atlas is to generate a model of the human proteome over the course of a cell cycle,” says Emma Lundberg from the KTH Royal Institute of Technology in Stockholm, the Chan Zuckerberg Biohub in San Francisco, and Stanford University.

The HCA database contains immunofluorescence images of several cell lines that were generated with a panel of around 16,000 antibodies. Analyzing the subcellular distribution of proteins in hundreds of thousands of images is a formidable challenge. Automated approaches exist, but they typically focus on a limited number of features and are not suitable for proteins that localize to multiple locations, which is a widespread phenomenon for human proteins. Lundberg was looking for alternative options, and her interest was sparked by a brief email from Attila Szantner from the start-up company Massively Multiplayer Online Science, based in Switzerland. Szantner suggested incorporating the image-classification task into a video game.

Lundberg considered Szantner’s suggestion “a great idea that immediately caught my attention.” Together, they developed it into a proposal for a mini video game, which they eventually pitched to different gaming companies. Project Discovery is now a minigame within EVE Online, which is played by half a million gamers. “Considering that it is three partners from different disciplines in three countries, [the project] worked amazingly well,” says Lundberg about the three-way collaboration.

The Project Discovery minigame can be played anytime and anywhere, and players can work on a small number of images to hundreds of images. The players go through a tutorial in which progressively more classes of protein localizations become available.



Project Discovery screenshot. Credit: Emma Lundberg

Lundberg says that it takes players about five minutes to become fully trained in classifying protein localizations into 30 different patterns. Players are rewarded for providing good outputs and for playing for a long time. Lundberg explains that players were initially rewarded when their chosen classification for an image matched the consensus of other players. However, the team quickly learned that players defaulted to the over-represented cytoplasmic localization class, which reduced accuracy. “Gamers tend to highly game the game,” says Lundberg. The team then introduced control images instead, for which the protein localization class was established, and rewarded the players only if they correctly assigned these images, which appeared randomly.

Because it was uncertain at the outset how well the approach of recruiting citizen scientists through an online game would work, Lundberg and her team also explored a deep learning approach for image classification. They trained their neural network, Loc-CAT, with images that had been annotated previously. Overall, the citizen science approach and the deep learning approach performed similarly, although there were differences in the details. For example, the machine learning approach performed better for common protein localization classes, provided substantial training data. Gamers, in contrast, tended to be better at recognizing rare localization patterns. The researchers exploited this difference by feeding the

gamers’ annotations into the deep neural network, which improved its performance.

At the time of publication, Project Discovery had resulted in the annotation of millions of subcellular protein localizations, and included patterns that the team had not previously annotated at a large scale. Lundberg hopes to eventually combine these individual protein localizations into a comprehensive model of the human proteome throughout the cell cycle.

In addition to the wealth of annotation data acquired through Project Discovery, the effort also provided interesting insight about how to engage a larger community in science projects. “I was very pleasantly surprised by the motivation of the gamers, by the feedback that we got, how many people said that they were genuinely happy to help science and to contribute to the greater good,” says Lundberg. Compared with platforms that pay for ‘microwork’, approaches like Project Discovery may result in better data quality, as contributors may be more engaged in the effort. “If you embed it into the narrative of the game, it can actually make people motivated in many different ways,” says Lundberg. People may be motivated for reasons as diverse as in-game rewards and a desire to help science or learn biology.

Despite the success of the Project Discovery approach, Lundberg sees further room for development. To this end, the team has set up a Kaggle competition to tap into the community of data scientists for help with improving automated tools for the image-classification task using the example of the HPA dataset.

Nina Vogt

Published online: 1 October 2018
<https://doi.org/10.1038/s41592-018-0162-4>

Research papers

Sullivan, D. P. et al. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* **36**, 820–828 (2018).