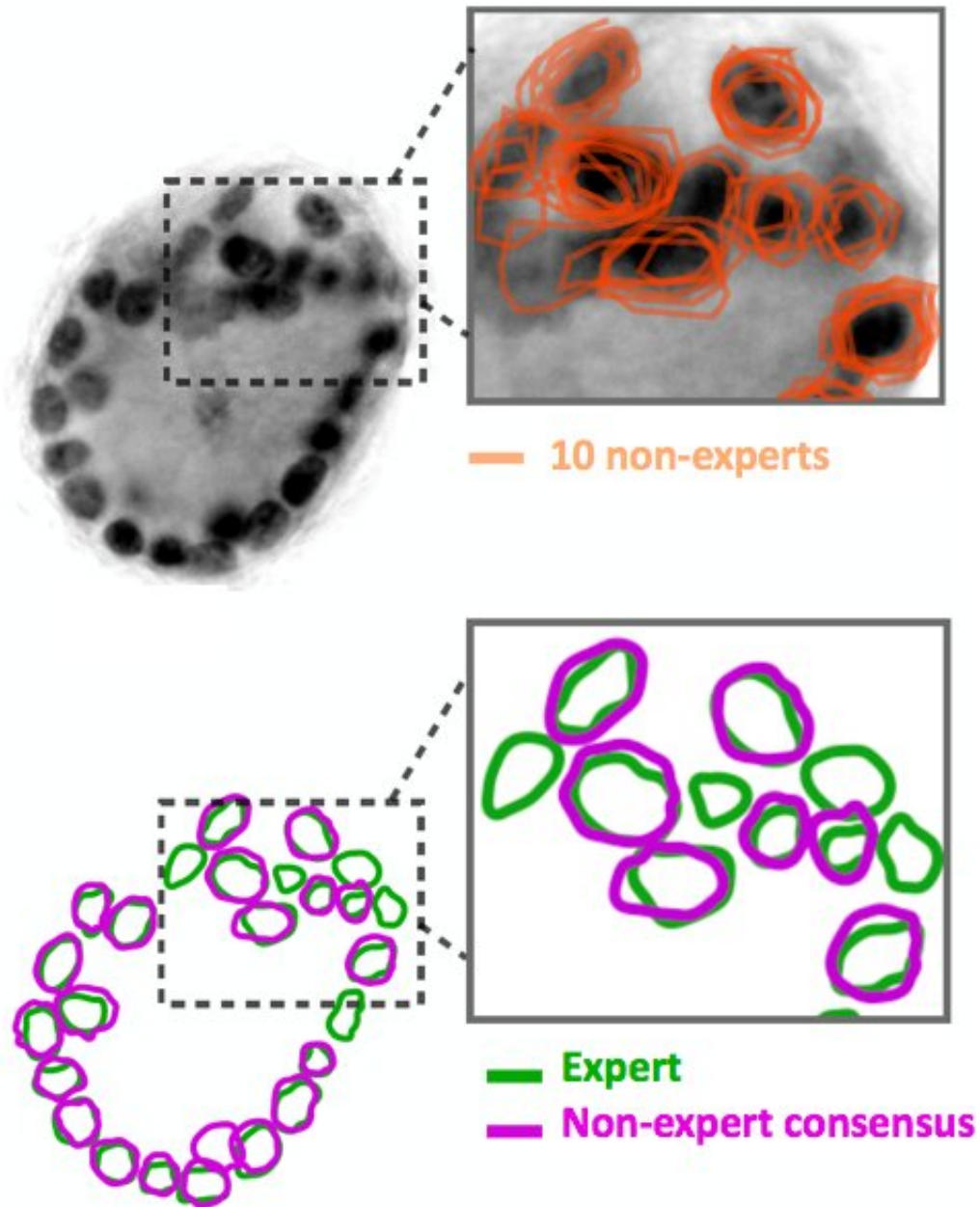


Teaching AI to learn from non-experts

August 1 2018, by Simone Bianco



Non-expert image annotations are noisy. Ten non-experts outlined the dark black

circles in the image, which are cell nuclei. Their results (shown in orange) do not match up exactly. Our algorithms are able to infer a consensus outline (shown in purple) from the noisy data. Compare this consensus with expert annotation of the same image (shown in green). Credit: IBM

Today my IBM team and my colleagues at the UCSF Gartner lab reported in *Nature Methods* an innovative approach to generating datasets from non-experts and using them for training in machine learning. Our approach is designed to enable AI systems to learn just as well from non-experts as they do from expert-generated training data. We developed a platform, called Quanti.us, that allows non-experts to analyze images (a common task in biomedical research) and create an annotated dataset. The platform is complemented by a set of algorithms specifically designed to interpret this kind of "noisy" and incomplete data correctly. Used together, these technologies can expand applications of machine learning in biomedical research.

Non-experts and noisy data

The limited availability of high-quality annotated datasets is a bottleneck in advancing [machine learning](#). By creating algorithms that can deliver accurate results from lower-quality annotations—and a system for rapidly collecting such data—we can help alleviate the bottleneck. Analyzing images for features of interest is a great example. Expert image annotation is accurate but time-consuming, and automated analysis techniques such as contrast-based segmentation and edge detection perform well under defined conditions but are sensitive to changes in experimental setup and can produce unreliable results.

Enter crowd-sourcing. Using Quanti.us, we obtained crowd-sourced image annotations 10–50 times faster than it would have taken a single

expert to analyze the same images. But, as one might expect, annotations from non-experts were noisy: some correctly identified a feature and others were off-target. We developed algorithms to process the noisy data, inferring the correct location of a feature from the aggregation of both on- and off-target hits. When we trained a deep convolutional regression network using the crowd-sourced dataset, it performed nearly as well as a network trained on expert annotations, with respect to precision and recall. Along with the paper describing our approach and strategy, we released the source code for our algorithm.

Applications in cellular engineering

Image analysis is central to many fields of quantitative biology and medicine. A few years ago we and our collaborators announced the NSF-funded Center for Cellular Construction (CCC), a science and technology center that is pioneering the new scientific discipline of cellular engineering. CCC facilitates close collaboration between experts of different disciplines, like machine learning, physics, computer science, cell and molecular biology, and genomics, to drive progress in cellular engineering. We aim to study and create cells that can be used as automated machines, or ad hoc sensors, to learn new and vital information about a variety of biological entities and their relationship with the environment they live in. We use [image analysis](#) to pinpoint the position and size of internal cell components. But even with advanced imaging techniques, exact inference of cellular substructures may be incredibly noisy, making it difficult to operate on the cell's components. Our technique can use this noisy data to correctly predict where the relevant cellular structures may be, allowing better identification of organelles involved in production of important chemicals or potential drug targets in a disease.

We believe our algorithms are an important first step toward more complex AI platforms. Such systems may use additional "human in the

loop" paradigms, by involving a biologist to correct mistakes during the training phase, for example, to further improve performance. We also see an opportunity to apply our method beyond biology to other fields where high-quality annotated datasets may be scarce.

More information: Alex J. Hughes et al. Quanti.us: a tool for rapid, flexible, crowd-based annotation of images, *Nature Methods* (2018).

[DOI: 10.1038/s41592-018-0069-0](https://doi.org/10.1038/s41592-018-0069-0)

This story is republished courtesy of IBM Research. Read the original story [here](#).

Provided by IBM

Citation: Teaching AI to learn from non-experts (2018, August 1) retrieved 6 May 2024 from <https://phys.org/news/2018-08-ai-non-experts.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.