Featured Prediction Competition

2018 Data Science Bowl

Find the nuclei in divergent images to advance medical discovery

$100,000 Prize Money

Booz Allen Hamilton · 739 teams · a year ago

Anne Carpenter

Annotation strategies used

posted in 2018 Data Science Bowl a year ago

5

Trust me, we debated long and hard about whether to reveal who was doing which annotations. I was outvoted so we kept it secret during the competition. So now it's time for the big reveal, for those who care:

We used two strategies for annotation; one is described here (literally outlining each nucleus with a tablet & pen):
https://www.kaggle.com/c/data-science-bowl-2018/discussion/48347

As far as I'm aware, all scored images in test stage 2 used this approach and were done by myself.

Images in training set and test stage 1 were a mixture: roughly 10% were done my myself using the strategy above; the remaining 90% were done by a team of people using the annotation tool that Allen made - he plans to make it public now, but doesn't want to support it going forward so don't pester him about it! It was some sort of super-pixel strategy.

Either way, each image was either annotated by or had annotations checked by a PhD biologist but generally speaking there was only one independent annotation done per image (we thought it better to have 2x as much data vs. having two annotators independently do the same images and provide a consensus labeling).