Title:

Review of "*Virus detection and identification in minutes using single-particle imaging and deep learning*" https://doi.org/10.1101/2020.10.13.20212035

Author:

Benjamin A. Himes, PhD
2020-Oct-16
Benjamin.Himes {AT} Umassmed <dot> edu

Note: The review presented here reflects my professional, but personal, reading of the manuscript and in no way reflects the view of my employer, the Howard Hughes Medical Institute, or my home institution, the University of Massachusetts school of medicine.

Scope:

While I am an expert in feature detection in very noisy microscope images, I do not work with fluorescence imaging. This is a rapid review and likely contains mistakes itself.

Summary:

The author's describe an assay combining TIRF, image segmentation and image classification to computationally sort and label virions present in solution. The benefit of such an assay could be widespread as it uses low-cost and non-specific reagents and obviates the professional intervention needed for RNA extraction and amplification in current gold-standard RT PCR test.

The authors use DNA oligos chemically fused to fluorescent dyes which bind non-specifically (from a Watson-Crick base pairing standpoint) to the virions in the sample. Image segmentation begins by using the available algorithms in MATLAB to pre-clean the data, removing regions that clearly have too much or too little signal. They rely on the more dependable signal from the red channel to do so. A final

selection criteria then requires the co-localization of the green channel to consider a region for further analysis.

Following segmentation, they have demonstrated the use of a CNN to classify these preselected regions from the field of view into “Virus” or “Not Virus.” And even SARS-Cov-2 positive or negative as demonstrated on clinical samples.

While encouraging, the results of the study do not fully support the conclusions drawn which are rife with hyperbole. While we all will benefit from fast and widespread testing, neither the scientific community nor the public on a whole benefit from sensationalism. The wide attention received due to the, imo irresponsible, promotion of a preliminary result by Oxford University prompted this review.

It is my hope that the authors find the following comments useful in strengthening their encouraging results and that this potentially impactful technical advance matures rapidly enough to benefit the current pandemic.

Major Issues:

- SARS-CoV-2 is highly pleomorphic, varying in shape, size, density of surface proteins and even the conformation of those structural proteins. Given that there is no clear understanding of what features the CNN uses in discriminating between viral types, the authors should have minimally presented an analysis of the types of virus used in the study and whether the “decoy” virions covered the range of features observed in SARS-CoV-2 in numerous cryogenic electron microscopy.

- There is no discussion on the choice of DNA oligos used in the labels. A quick search with the sequence (just in Google, not Blast) shows that they have been used in previous studies. Why did the author’s stick with oligo used in the green channel given its dismal precision (False positive rate?) Did they try any other oligos?

- The selection criteria in the initial segmentation should be explained more clearly. Throughout the paper, pixel units should be presented as length units.

(Especially as physicists!) The physical size of SARS-CoV-2 is on average ~120 nm which is surely more than 10x a single fluorophore/oligo label. Perhaps this is common in TIRF studies, but some explanation of the relationship between signal dispersion and physical size of the bound object is required.

- The labelling of the mixed samples for training is central to the success of the CNN classification. As such it needs to be clarified in the text:
    - For the lab grown samples, it seems like the combination of different virions only happens after segmentation *in silico*. If this is the case it should be briefly described and if not, the method for defining the ground truth should be included.
    - For the clinical samples, it is implied that each sample is tested for both SARS-CoV-2 and the other viruses listed, but this is not stated explicitly. The viruses present in each mixed clinical sample used to quantify SARS-CoV-2 precision and recall should be presented either tabularly or graphically.

- The following claim is not supported by any evidence and should be qualified as such in the discussion.
    - *"The decrease in accuracy (compared to the laboratory-grown viruses) reflected the greater heterogeneity and complexity of clinical samples (e.g., varied storage conditions, wide range of virus concentrations, presence of residual cellular material, different sampling techniques)."*

- As it is again unclear how the clinical sample's viral compositions were quantified, it is hard to parse the quote below. It seems plausible that a patient could have both the flu and a seasonal hCoV. What then would be the false positive rate? Is it the product of the two probabilities, i.e. 51% or just chance? Perhaps more troubling is that they seem to be referring to accuracy in the validation data set, and nowhere present any comment on whether that accuracy may be subject to bias or overfitting.

- *"In spite of these issues, the network could also distinguish SARS-CoV-2 from seasonal hCoVs with a validation accuracy of ~73% (Fig.3C), and SARS-CoV-2 from Flu A with a validation accuracy of ~70% (Sup.Fig.6B)"*

- In the following statement, where the authors use their test set, which the network hasn't seen and should not be overfit (see previous comment.) How much "more than" 50% are correctly classified? What happened in the samples that are excluded from these statistics and why were they excluded?

    - *"The network was able to classify more than 50% of BBXs correctly in 8 of 10 samples tested for seasonal hCoVs vs. negative, and in 8 of 9 samples tested for SARS-CoV-2 vs. hCoV; results can be further improved by increasing the number of samples used for training."*

- The authors only tested a few strains of a few viruses. They did not present any results with alternate labels, immobilization techniques or imaging strategies. The following statement is therefore, quite egregious.

    - *"Our algorithms are extremely versatile and can be trained to differentiate between many different viruses, independently of how they are labelled, immobilized and imaged.*"

Minor Issues:

- Throughout, suggest that "DNAs" should read "DNA oligonucleotides" or "DNA oligos" to be clear you are not referring back to viral genomic material. (Yes SARs is and RNA virus, but presumably the test could be used with DNA viruses present)

- The authors might define what they mean by semi-major-to-semi-minor-axis-ratio. Additionally, this and the other statistical descriptors may be used by the CNN to classify, but they may not. This might be mentioned.

- The limit of detection was measured, but there is no discussion on how this relates to current state of the art and how that relates back to catching active infections in patients.

- This one is picky, but the labelling takes seconds. It is not "*instantaneous*."

- It is unclear what the authors think the advantages are or why those are justified.
    - "*The non-specific detection of intact viral particles (rather than genome fragments) can report directly on infectivity, and has the advantages"*
- Pixels should be changed or augmented by length.

- There is no discussion or justification for transforming from a softmax output (relative probability of class assignment) to a binary classifier. Is this so their binomial PDF can be used to estimate the noise floor?

- Equation 3 is missing FP in the denominator.