

OCR Handwriting Project Special Report

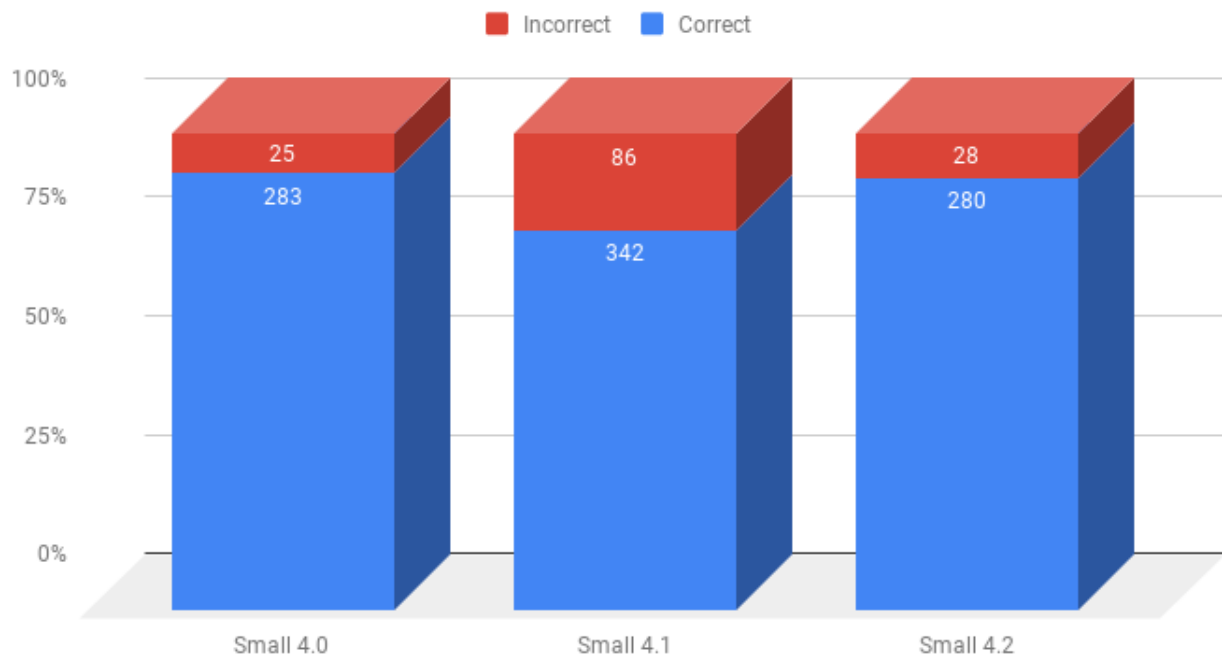
Matthew Mulhall

matthew.l.mulhall@uconn.edu

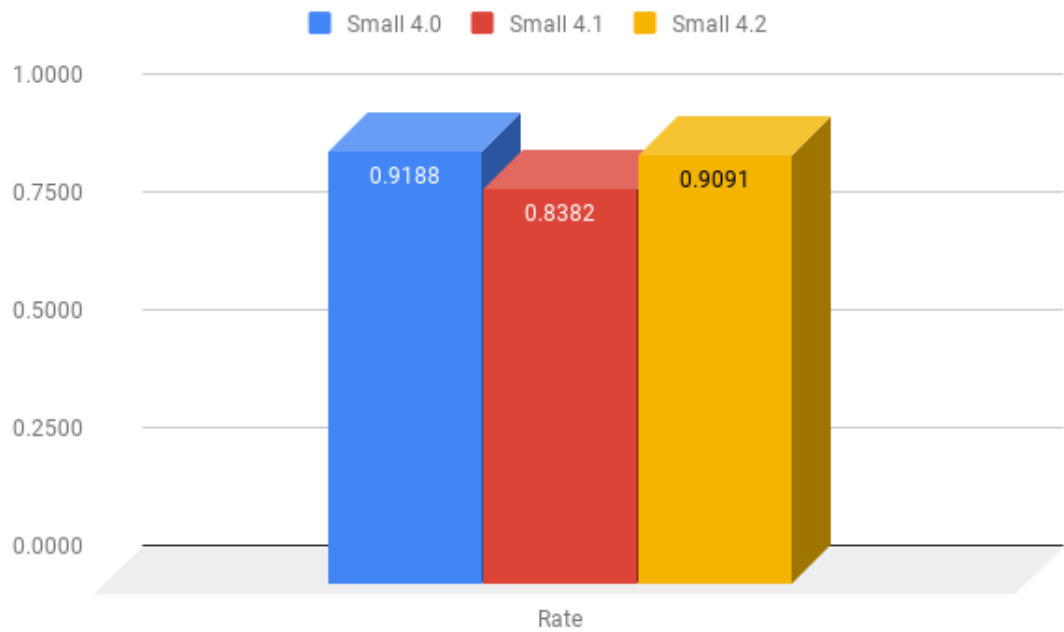
July 31, 2019

- (i) Today's tests were run in order to decide the future of the project as far as trajectory. After several tests I think that it is clear that our data set must be increased. If I had to give an estimation I would say that to realistically solve this problem at high success rates, we would need something on the order of 3-5x more images per sample. I believe this is due to the inter-class noise that our model is picking up. It is having a hard time deciding between certain letters because their characteristics are not defined different enough through the data. That being said we still have great success rates given the reality of this problem. For example, our 19 class test currently runs around 80% accurate. To put this into perspective, it is still 15.2 times better than random guessing. To me, this means our model is built correctly, yet it still doesn't have enough data to properly discern classes. Below I have shown the conclusions of the tests to support my hypothesis.

Small 4.0, Small 4.1 and Small 4.2

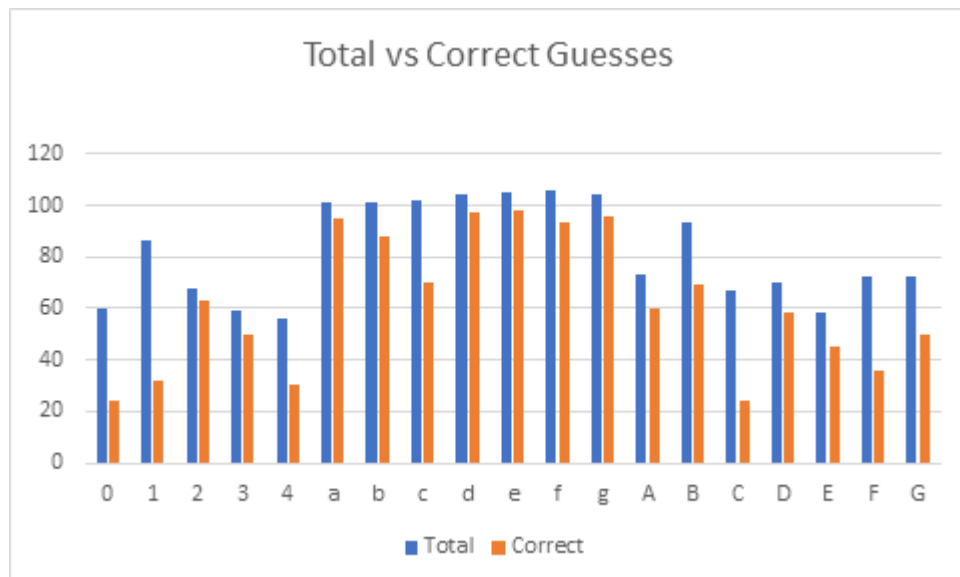


(ii)



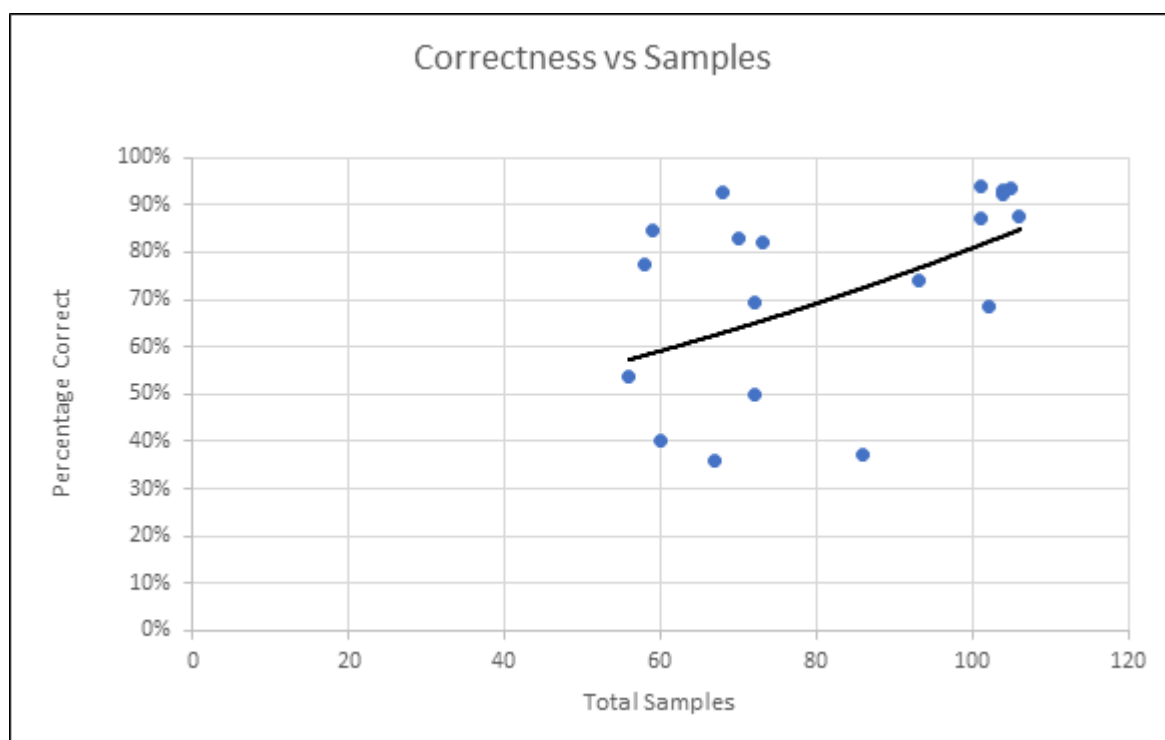
(iii)

- (iv) Notes: Small 4.0 is the base neural network for 4.1 and 4.2. The differences are as follows:
- (v) 4.0 vs 4.1: 4.1 increased the augmentation from 250 photos per class to 500 photos per class, which is reflected in the increase in overall numbers for 4.1. This was not a positive improvement.
- (vi) 4.1 vs 4.2: 4.2 retained the same model as 4.1 and 4.0. 4.2 Used a augmentation size of 300 as well as removed the horizontal flip option from the augmentation. I did this as I thought that flipping a d or b horizontally could lead to classification issues. It does not seem that this had a positive effect.
- (vii) I also compiled some useful data regarding to the medium (19 class) set.
- (viii) Amendment 7/31/2019. 22 class classification yielded a test result of 86.4%. This model used a data set that was heavily expanded by 800 over the original samples, and had a dense layer with 2,048 nodes. I also removed one of the convolution layers, which seems to have yielded good results. This is a 6.4% increase over the previous best, and with a larger classification task. Compared to the next best (15.2 times better than guessing) this test yielded results that were 19 times better than guessing on a 22 way classification problem.
- (ix) The following is a graph showing the total number of samples vs correct guesses that was achieved during a test on the 7th iteration of the medium set model.



(x)

- (xi) The following graphic used the previous graphs data to find the relationship between sample sizes and correctness.



- (xii)
- (xiii) Given the conclusions of the tests as well as plotting the information into graphs, the model needs more data. The previous chart's trendline clearly suggests that there is a positive relationship between samples taken and the model's success. Although this seems obvious, it was important to isolate each character, as well as rule out the interaction of certain characters. I.e. it was important to figure out if 'b' and 'd' interacted, or if they had high success despite their similarity.
- (xiv) New best model data:

(xv)

