Daily Log

Monday September 23

Ranked writers based off of how many forms they contributed to the IAM database. I created a dictionary mapping a writer to a list of the forms he contributed to the database by going through IAM's forms.txt file, which lists each form along with information about the author, the number of sentences, and word segmentation. For our purposes, we are only interested in the author and the form ID. Then I used python's sorted function to convert the dictionary into a list of tuples containing each element's key and value in increasing order of value length, in this case the number of forms contributed. I then reversed the list so it would be in decreasing order by number of contributions.

Tuesday September 24

Generated the dataset by implementing the image patching method on the top 20 writers with the most contributions. For each of the first 20 writers in the list from yesterday, I go through each form that they contributed and run the image segmenting method, which makes image patches from each line of a given form and saves them in the data folder within another folder named after the author, so the data will be organized by writer. I also worked on setting up a GitHub repository so my partner and I can share our code more effectively and easily, and in the last few minutes of class, I started thinking about how I was going to install Keras and Tensorflow.
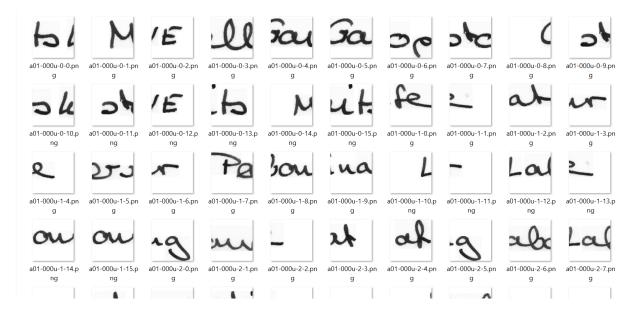
Thursday September 26

Today I worked on getting Keras and Tensorflow installed so I can start working on the deep-learning network. I ran into some issues getting Tensorflow installed, since I was using python-3.7 32-bit. After looking into it a bit more, I found out that python-3.6 is the most recent version of python that Tensorflow is compatible with and that Tensorflow only works with 64-bit versions of python on Windows. I created a virtual environment for my project in PyCharm that uses python-3.6 64-bit and was able to install Tensorflow there.

Timeline

Date	Goal	Met
September 16 - 20	Finish proposal and have a defined plan on how to generate sample image patches.	Yes, proposal was turned in and the plan was finished
September 23 - 27	Implement plan for generating sample image patches.	Yes, we were able to generate the sample image patches
September 23 - 27	Get organized data for top 20 writers	Yes, created data folder containing folders for each writer with the image patches we cropped from their writings.
September 30 - October 4	Feed the data into a CNN	
October 7 - 11	Improve our deep-learning network	

Reflection

This week, I organized the data for the top 20 writers in a folder called data, which contains folders for each of the top 20 writers containing the sampled image patches from each line of each of their forms. Once I had organized the data, I started working on downloading Keras so I could feed the

data into a deep-learning network. In order to use Keras, I needed to get Tensorflow first, which unfortunately does not support python 3.7 or any 32-bit version of python, which was an issue for me since I have been using python 3.7 32-bit. I knew I needed to use a virtual environment for this project if I wanted other stuff that I have on my computer that works with the version of python I originally had to keep working, but I didn't really know much about using a virtual environment, so I had to do some research. Thankfully, pyCharm, the IDE I am using, makes it really simple to use a virtual environment, so once I figured out how to set everything up and install packages on it, everything worked out fine. However, I did end up having to spend an entire period on a task that I had thought would only take a minute or two.