Processing Handwritten Documents

Alex Raita, Haris Sheikh

Project Idea

We want to attempt to take handwritten documents and convert them to formatted text. That means recognizing letters, words, spaces, indents, and line breaks in order to piece together text formatted in same way as the document. After this piece of the project is finished, we want to also be able to recognize mathematical equations (not just mathematical symbols) and be able to solve them, as well as output them in a nicely formatted way (i.e. LaTeX).

We anticipate that we will have difficulties in determining the order words and letters come in, as well as determining the position of different punctuation. For example, sometimes in a handwritten paragraph, it is hard to tell if something is an apostrophe or a comma since it isn't obvious which line the apostrophe or comma belongs too. Also, in mathematical equations we have to be able to identify relations between symbols (like exponent and fractions).

Workload

Although it is a relatively simple task to identify letters, it is a much more difficult task to identify words, sentences, and punctuation. We think it will take quite a bit of effort to optimize our detection efforts in order to be accurate enough that we can read full paragraphs completely, along with mathematical symbols. Here are some proposed steps:

- Identify locations of letters and/or words. Letters can be detected using techniques from class, while identifying locations of words can be done using machine learning models. We can reference here for word detection
- Next, we want to be able to recognize the individual letters and map them. This is a well-known problem with many available datasets to train a machine learning model on. Example of mathematical symbols, example of handwritten characters.
    - If we end up working with entire words, then we need to recognize the letters inside the individual words using the same techniques as above
- Piece together the recognized letters in to words and the words into sentences. This requires us to be able to process the locations of words and letters in order to make decisions about what constitutes a word and a sentence.
- Output formatted text with spaces (and punctuation) so that it closely matches the format of the document
- Detect if a 'word' is a math equation. This can mostly likely be done by recognizing numbers and math symbols in a word and deciding if they are an equation.
- Detect the symbols in an equation and use relative positions of these symbols to output a formatted equation
    - We can solve said equation by parsing via the shunting-yard algorithm and then evaluating the resulting Reverse Polish notation expression
    - Alternatively we can use Wolfram Alpha's public API to evaluate these equations