

Text Recognition

Choice of Dataset:

We will be using the [NIST Special Database 19](#) dataset (If this one proves too difficult to use, we might use the [EMNIST Dataset](#) which is derived from NIST and is standardized to 28x28 like the original MNIST Dataset) to first train a model capable of distinguishing between words, then we may use our class notes or other datasets to train the other model to break a word image into letters.

Methodology:

Our model will consist in a sequence of three convolutional neural networks. The first, will receive a preprocessed image from an algorithm that scales the image to fit the model correctly and have the goal to correctly separate words from the image. Every new image will become the input of the second CNN and be divided into letters. Letters will be the input of the last CNN and be identified. An algorithm will assemble all letters into words, and all words into sentences. Therefore, our whole model will receive images of text and output its text typed.

We are unaware of how loss could be quantified in terms of computer vision, i.e. the first two CNNs, but we plan to use categorical cross-entropy loss for the last CNN while the last would have a softmax activation function in its last layer.

Application:

We plan to create a web app or even possibly a mobile app, where users would take a picture of text, send it directly to the application, and receive a copyable text output. Depending on how challenging it is, we also plan to allow users to submit pictures of simple math problems as input, process the image with the pre-described CNNs and then use *SymPy* to compute the answer.