Project Plan Report
Optical Handwritten Words Recognition
(Brandon Rudolph, Patrick Rippl, Yuankai Wang, Chen Yin)

Our **minimum goal** is recognizing groups of identified characters as words. Our **reasonable goal** is recognizing a sentence without cursive writing. Our **stretch goal** is to recognize a whole paragraph.

We have found three different types of datasets, which are **single letter train set**, **simple words data set**, and a **test set with more words, sentences and paragraphs**. We also have an advanced dataset, which contains the whole paragraph for our stretch goal.

Before feeding the test set into the net, we will apply a few simple morphology methods on words to **clear the noise** and isolate each letter. After this **preprocessing**, we will **separate** each letter in the word by using the MATLAB $regionprops$ function and find the bounding box of each letter. In our training part, we will feed **62 different classes** (0-9, A-Z, a-z) of character to our net and we will use **AlexNet** as preliminary classifier.
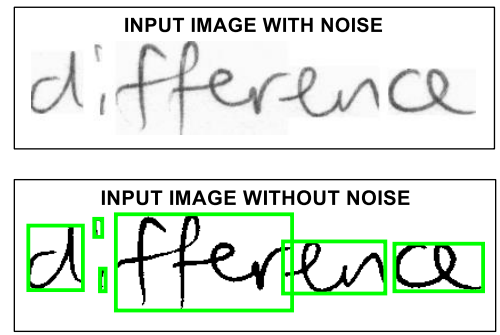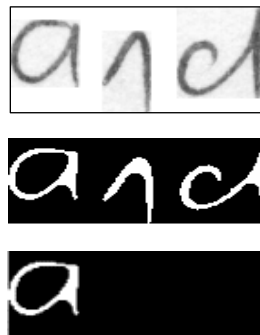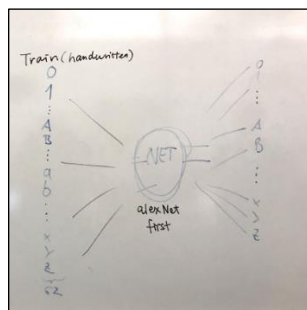

Figure 1 Some of the current code

Figure 2 Exemplary results for easy ('and') and more difficult words ('difference')

The letters i and j have small dots that are separated from the main part of the letter. When preprocessing images to be fed to the network **the small dot will be ignored** to keep the processing simple. Meanwhile, we cannot recognize cursive writing because we cannot segment each letter correctly.
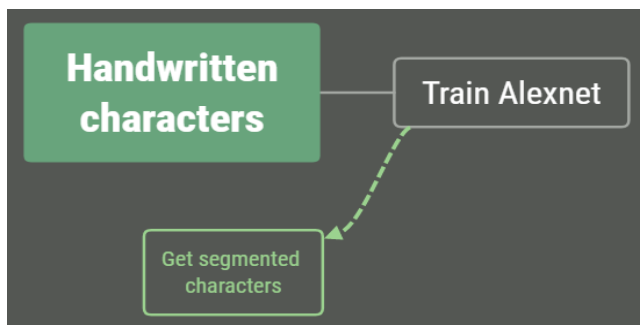

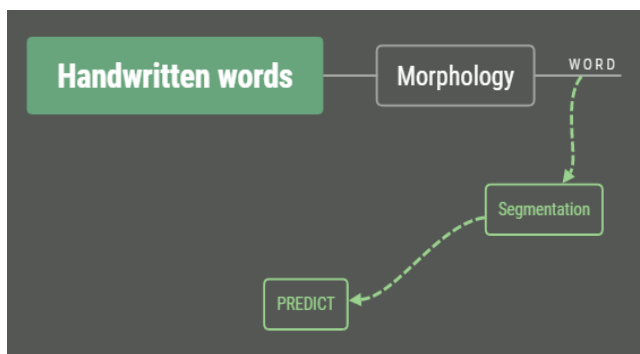*Figure 3 Part I: Train our net with handwritten character set (train data set)*

Figure 6 Part II: Preprocess words from test set and predict them with our trained net

Goal	Deadline
Suitable dataset (characters & words)	01/28
Run AlexNet on Dataset and get preliminary results and set a benchmark	02/02
Have a morphology function for test data set	02/02
Separate individual letters within words	02/02
Identify individual handwritten characters	02/02
Identify groups of characters as words	02/09
Output identified, handwritten words as typeset words	02/09
Convert series of handwritten words (e.g. sentences, paragraphs, etc.) into a typeset sentence/paragraph	02/09
Finish report and prepare presentation	02/11
Extend code (improvements, more words, etc.)	- finish

Figure 4 above: Deadline scheduling

Figure 5 right: Tasks until 02/02