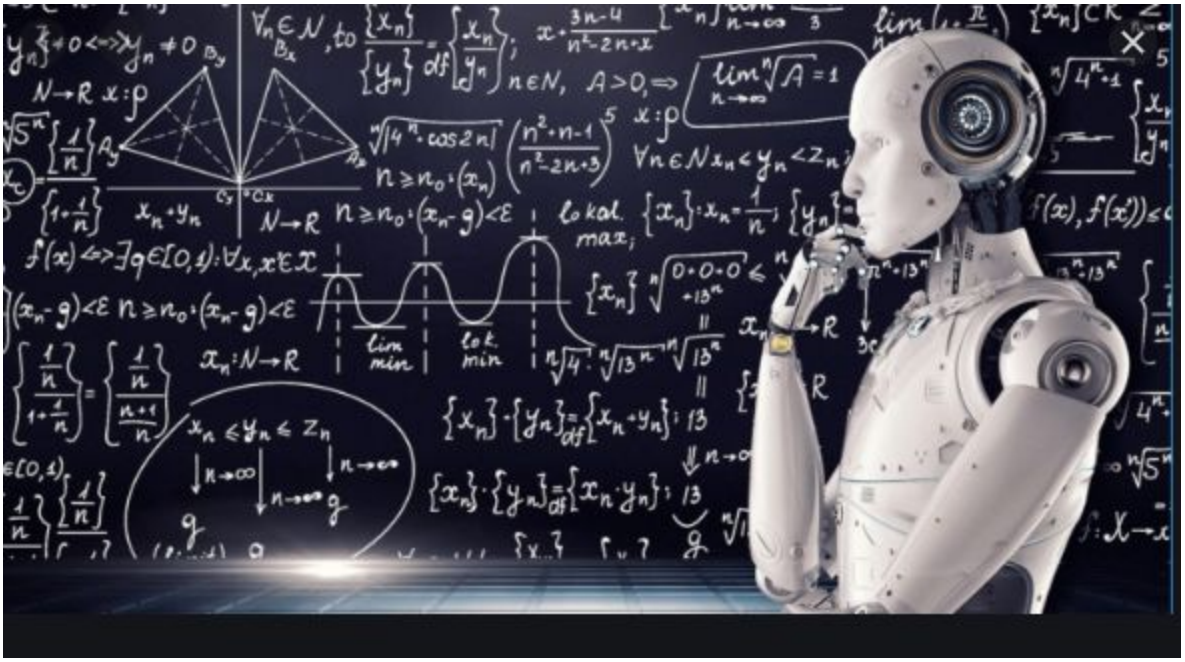
Project Proposal

Handwritten Text Recognition System



Team

Ashraqat Ahmed Sheta (13)
Fatma Ibrahim Hemeda (42)

Problem Statement

Handwritten Text Recognition is a technology that is much needed in the world today. There are many documents that are written in papers by hand and now with computers everywhere in the world we need to access these documents with computers which is a very difficult process to do as we can't rewrite them again in computers so this technology can definitely be used to do this task . Thanks to deep learning techniques the error percentage is very small and acceptable .
The task will be done by entering a handwritten document to the network and the document will be computerized .
And this is not the only usage of the technology as people can use it to enter their prescriptions which are in general not understandable and when they are computerized , they can easily read the content .
Writers and traditional peoples who still prefer to use a pen for writing and need other peoples to work for them to rewrite their papers as a softcopy can also benefit from this technology.

Current state of the art accuracy for the handwritten text recognition problem:

There are many models that were used to solve this problem and each has a different accuracy .
An average accuracy is between 75% and 85% the accuracy of the same model may differ from one language to another and also can differ between numbers and letters (the accuracy of numbers recognition is higher than letters recognition .
Cursive handwriting has less accuracy as it is difficult to interpret handwriting with no distinct separation between characters.
There are models that exceeds this accuracy :

- An efficient approach towards the development of handwritten text recognition systems is 3-layer Artificial Neural Network (ANN). The feature vectors are first pre-processed in order to remove the noise and then applied to the ANN along with the generated target vectors; that are generated on the basis on input samples. 55 samples of each English alphabet are used as a ANN training process in order to make sure the general applicability of the system towards new inputs. Two different learning algorithms are used in this approaches ResilientBack-propagation or Scaled conjugate gradient Additive image processing algorithms are also developed in order to deal with the multiple characters input in a single image, tilt image and rotated image. The trained system provides an average accuracy of more than 95 % with the unseen test image.
- In 2018, researchers applied the technique of DCNN (deep CNN) for recognizing the offline and handwritten Arabic characters . An accuracy of 98.86% was achieved .
- **Deep learning enables recognition of text to 99.73% accuracy:**this is the best accuracy reached although there is no mentioned model to prove this **.**

A short survey of available models and solutions for the handwritten text recognition problem

Many recognition studies have been made for offline and online handwritten characters of major languages used worldwide like English, Chinese and Indian.but they all suffer with some sort of drawback like low conversion speed, low accuracy, higher false detection rate and poor performance with noisy input etc. Thus, recognition studies of handwritten character image samples still remain relevant because of their enormous application potentials.

These are some of the models that are used to solve the handwritten text recognition problem:

- A model is proposed which uses different types of neural networks. It consists of convolutional NN (CNN) layers, recurrent NN (RNN) layers and a final Connectionist Temporal Classification (CTC) layer
- Rajib proposed a handwritten English character recognition system based on the Hidden Markov Model (HMM).This method made use of two different feature extractions namely global and local feature extraction. Global feature extraction includes many features like gradient features, projection features and curvature features in the numbers of four, six and four respectively. Whereas local features are calculated by dividing the sample image into nine equal blocks. Gradient feature of each block is calculated using four features vector, which makes the total number of local features as 36. This resulted in fifty features (local + global) for each sample image. Then, these features are fed into HMM model in order to train it. Data post processing is also utilized by this method in order to decrease the cross classification of different classes. This method takes a lot of time in training and feature extraction. Moreover, it performs poorly in case of such inputs,when many characters are combined in a single image.
- Velappa Ganapathy proposed a recognition method based on multi scale neural network training. In order to improve the accuracy, this method used a selective threshold, which is calculated based on minimum distance technique. This method also involves the development of GUI, which can find out the character throughout the scanned image. This method provides an accuracy of 85% with moderate level of training. This method used large resolution images (20 × 28 pixels) for training with lesser training time.
- T. Som used fuzzy membership functions to improve the accuracy of the handwritten text recognition system. In this method,text images are normalized to 20 × 10 pixels and then a fuzzy approach is used in each class. Bounding box is created around the character in order to determine the vertical and horizontal projection of the text. Once the image is cropped to a bounding box, it is re-scaled to the size of 10×10 pixels. Then, cropped images are thinned by the help of thinning operation. In order to create the test matrix, all these pre-processed images are placed into a single matrix; one after another. When new (test) images are presented by the user, it is tested for the matching against the test matrix. The method was fast but it provides a low accuracy.
- Rakesh Kumar proposed a method in order to reduce the training time of the system by utilizing a single layer neural network. Segmented characters are scaled to 80 X 80 pixels. Data normalization is performed on the input matrices to improve the training performance. But their result has a low accuracy rate.
- By using the Euler number approach, speed and accuracy are improved. Many preprocessing like Thresholding, thinning and filtering operations are performed on the input image so that cross error rate can be minimized. Three techniques are utilized for better segmentation. After segmentation, the input image is resized to the size of 90 X60 pixels. Then after the Euler number is calculated for each text and then they are divided into 54 zones, such that each contains 10×10 pixels. The average value of

each zone (row and column wise) is used as the feature vector of the character. These features are fed in to a feed forward back propagation neural network (FF-BP-NN), which have a configuration of 69-100-100-26. This system classified the data into 26 different English letters. This method performs well but it does not include the classification of small English letters.

- Anshul and Mehta proposed their work based on the heuristic segmentation algorithm. Their system performs identification of valid segmentation points between handwritten letters quite well. Fourier descriptors are used in this approach for feature extraction. After a successful segmentation, Discrete Fourier Coefficients are calculated (a[k] and b[k]) for the input image. Here k varies from zero to (L-1) and L represents the boundary points of input image.This method tried to provide classification of a total 52 characters (26 uppercase English letters and 26 lower case English letters).It also provides a comparative analysis of different classification methods. The method also incorporates post processing in order to reduce the error rate but it suffers with a low accuracy and high cross classification rate; because of the non-optimal choice of features.

- Serrano proposed a novel interactive approach for handwritten character recognition. The system requires human suggestion for only those inputs for which the system gets confuse.Although It keeps the accuracy to a high level, it increases the human lead. The only problem was that the system was not fully automatic and requires human intervention for operation

A detailed description of the model to be used

We will use a neural network model which consists of 5 layers of convolutional NN (CNN) layers, 2 layers of recurrent NN (RNN) layers and a final Connectionist Temporal Classification (CTC) layer.

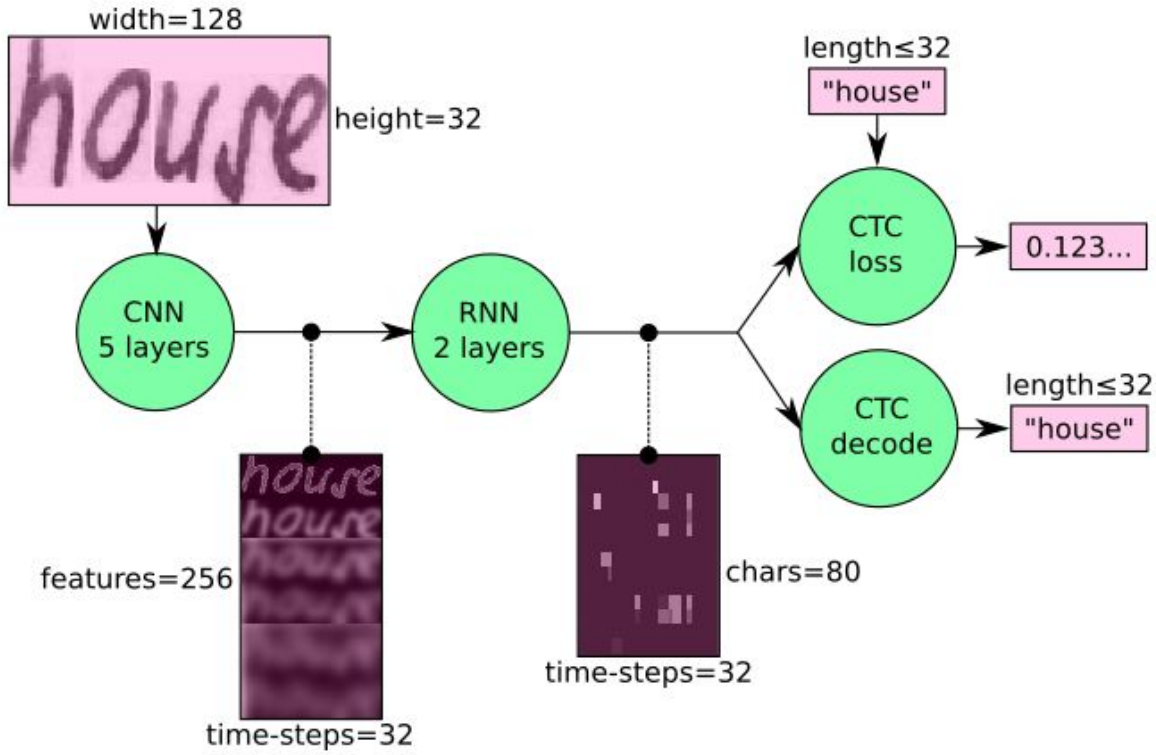


Fig. 2: Overview of the NN operations (green) and the data flow through the NN (pink).

the text is recognized on character-level, therefore words or texts not contained in the training data can be recognized too.

CNN: the input image is fed into the CNN layers. These layers are trained to extract relevant features from the image. Each layer consists of three operations. First, the convolution operation, which applies a filter kernel of size 5×5 in the first two layers and 3×3 in the last three layers to the input. Then, the non-linear RELU function is applied. Finally, a pooling layer summarizes image regions and outputs a downsized version of the input. While the image height is downsized by 2 in each layer, feature maps (channels) are added, so that the output feature map (or sequence) has a size of 32×256.

RNN: the feature sequence contains 256 features per time-step, the RNN propagates relevant information through this sequence. The popular Long Short-Term Memory (LSTM) implementation of RNNs is used, as it is able to propagate information through longer distances and provides more robust training-characteristics than vanilla RNN. The RNN output sequence is mapped to a matrix of size 32×80.

CTC: while training the NN, the CTC is given the RNN output matrix and the ground truth text and it computes the **loss value**. While inferring, the CTC is only given the matrix and it decodes it into the **final text**. Both the ground truth text and the recognized text can be at most 32 characters long.

Input: it is a gray-value image of size 128×32. Usually, the images from the dataset do not have exactly this size, therefore we resize it (without distortion) until it either has a width of 128 or a height of 32. Then, we copy the image into a (white) target image of size 128×32. This process is shown in Fig. 3. Finally, we normalize the gray-values of the image which simplifies the task for the NN. Data augmentation can easily be integrated by copying the image to random positions instead of aligning it to the left or by randomly resizing the image.

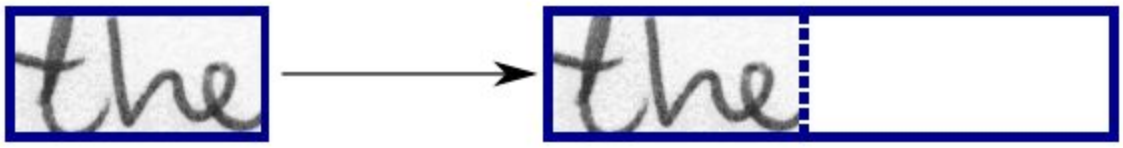


Fig. 3: Left: an image from the dataset with an arbitrary size. It is scaled to fit the target image of size 128×32, the empty part of the target image is filled with white color.

Why use this model ?

- NN-training is feasible on the CPU as the use of cpu is much better from GPU in the availability of resources
- The model uses CNN and RNN which will be good practise on which we learn in the class and we will understand what we are dealing with.
- This model is deeping in deep learning but other models depend on Artificial intelligence and machine learning or algorithms and deep learning can have better accuracy .

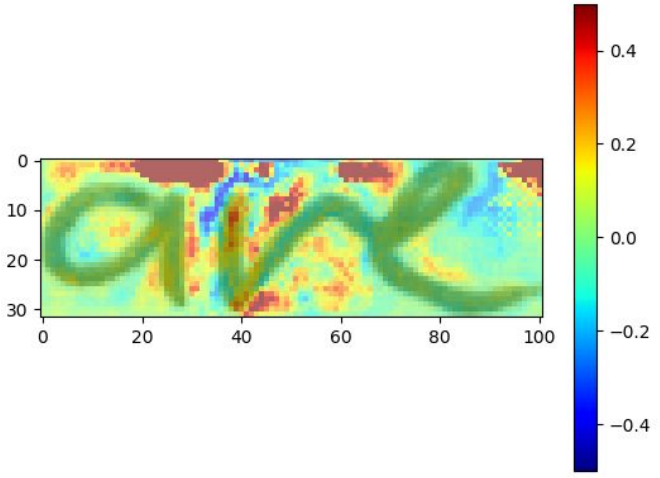
The proposed updates to the literature model

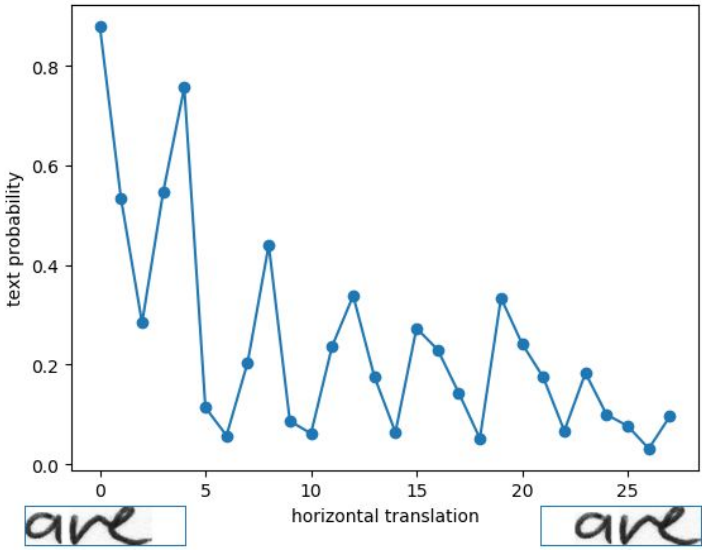
Data augmentation: increase dataset-size by applying further (random) transformations to the input images. At the moment, only random distortions are performed.
The benefit of this is to improve accuracy,improve the translation invariance of the model.
To make training not overfit and that improves the model.

Write about how you will evaluate your results, what datasets you will use, what kind of evaluation metric you will use to compare your results, and what types of plots/graphs will be used to point out the comparison results.

The CTC layer either calculates the loss value given the matrix and the ground-truth text (when training), or it decodes the matrix to the final text with best path decoding or beam search decoding (when inferring)
Results are shown in the plots below. The pixel relevance (left) shows how a pixel influences the score for the correct class. Red pixels vote for the correct class, while blue pixels vote against the correct class. It can be seen that the white space above vertical lines in images is important for the classifier to decide against the "i" character with its superscript dot. Draw a dot above the "a" (red region in plot) and you will get "aive" instead of "are".

The second plot (right) shows how the probability of the ground-truth text changes when the text is shifted to the right. As can be seen, the model is not translation invariant, as all training images from IAM are left-aligned. Adding data augmentation which uses random text-alignments can improve the translation invariance of the model. More information can be found in





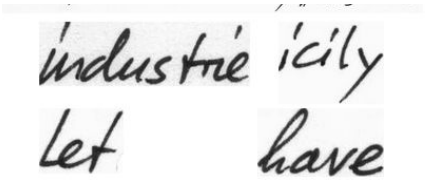
A survey of available datasets for your course project problem.

The IAM Handwriting Database contains forms of handwritten English text which can be used to train and test handwritten text recognizers and to perform writer identification and verification experiments.
The database contains forms of unconstrained handwritten text, which were scanned at a resolution of 300dpi and saved as PNG images with 256 gray levels.

Characteristics: the IAM Handwriting database is structured as follows:
657 writers contributed samples of their handwriting,1'539 pages of scanned text

5'685 isolated and labeled sentences, 13'353 isolated and labeled text lines, 115'320 isolated and labeled words

A detailed description of the dataset to be used. Also mention why will you use this particular dataset.



This dataset contain forms of a extracted words as shown. Because the model is used for extracted words not line text.

Graduation project brief problem statement

The work force is the basic element of any corporation that if correctly analyzed and managed accurately can lead to prosperity of this entity and revolutionary predictions that may save resources or achieve massive profits if analytics were accurate enough to read the future.

The amount of data is massive and highly variant and distributed over many databases that are not connected or mapped in a clear way that helps extracting valuable information , in addition the scatter of this data makes it more difficult for HR employees to analyze them for certain purposes which waste unnecessary time and analysis might not be very accurate which might lead latter on to taking wrong decisions that would cost the corporation an unnecessary cost whether as using unneeded resources (recruiting , time consuming ,…) or waste of already existing resources (an excellent employee skills unused or a project with high profit gets rejected for wrong predictions based on inaccurate analytics.

This data must be gathered in a Data-WareHouse that connects all variant aspects of the corporation where it can be managed dynamically viewed, updated and analyzed applying data mining techniques with trusted predictions for better decision making and less HR employees.

Resources and papers

https://www.ijeat.org/wp-content/uploads/papers/v8i3S/C11730283S19.pdf

http://ijcsit.com/docs/Volume%207/vol7issue1/ijcsit2016070101.pdf

https://www.researchgate.net/profile/Hazem_El-Bakry/publication/298808334_Handwritten_Text_Recognition_System_based_on_Neural_Network/links/56f2ac6d08ae0c8aa1d032ab/Handwritten-Text-Recognition-System-based-on-Neural-Network.pdf?origin=publication_detail

https://www.researchgate.net/publication/321029318_Investigation_on_Deep_Learning_for_Off-line_Handwritten_Arabic_Character_Recognition

https://towardsdatascience.com/https-medium-com-rachelwiles-have-we-solved-the-problem-of-handwriting-recognition-712e279f373b

https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5

http://www.fki.inf.unibe.ch/databases/iam-handwriting-database