Handwritten Character Recognition Using CNN

Author[1:] SHODA BHAVYA,
Author[2]: SHAIK CHANDINI,
Author [3]: SUVARNAGANTI DEEPIKA,
Author[4]: PARVATHANENI KIRANMAYI.

[1] *Student, ECE, VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY, A.P., INDIA*
[2] *Student, ECE, VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY, A.P., INDIA*
[3] *Student, ECE, VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY, A.P., INDIA*
[4] *Student, ECE, VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY, A.P., INDIA*

ABSTRACT

This paper seeks to classify an individual handwritten word so that handwritten text can be translated to a digital form. The problem is a formulation of both classification as well as sequence modeling due to its inherent nature that the occurrence of individual characters depends on the previous few characters. We investigate various neural network architectures to build a model that performs character segmentation, and then reconstruct each word. We will train and test our model on IAM handwriting database.

Keyword: - *Key word1*: *Character recognition, Key word2: convolution neural networks, Key word3: IAM database, Key word4: sequence modeling key word5: segmentation.*

---

1. INTRODUCTION

Most of the documents and thesis were written in journals, books in olden days. It is very difficult to store, access these notes. To share and search in these notes is difficult. There is a chance of losing this information is very high when some calamities occurred. The main problem encountered when dealing with handwritten English characters is that characters written by different persons representing the same character are not identical but can vary in both size and shape. The fast variation in personal writing styles and differences in one person's writing style depending on the context is another problem encountered when trying to recognize English handwritten characters. Handwritten character recognition explores the task of classifying handwritten text and to convert into digital format.

2. RELATED WORK

The aim of this project is to further explore the task of classifying handwritten text and to convert handwritten text into the digital format. Handwritten text is a very general term, and we wanted to narrow down the scope of the project by specifying the meaning of handwritten text for our purposes. In this project, we took on the challenge of classifying the image of any handwritten word, which might be of the form of cursive or block writing. This project can be combined with algorithms that segment the word images in a given line image, which can in turn be combined with algorithms that segment the line images in a given image of a whole handwritten page. With these added layers, our project can take the form of a deliverable that would be used by an end user, and would be a fully functional model that would help the user solve the problem of converting handwritten documents into digital format, by prompting the user to take a picture of a page of notes. Note that even though there needs to be some added layers on top of our model to create a fully functional deliverable for an end user, we believe that the most interesting and challenging part of this problem is the classification part, which is why we decided to tackle that instead of segmentation of lines into words, documents into lines, etc.

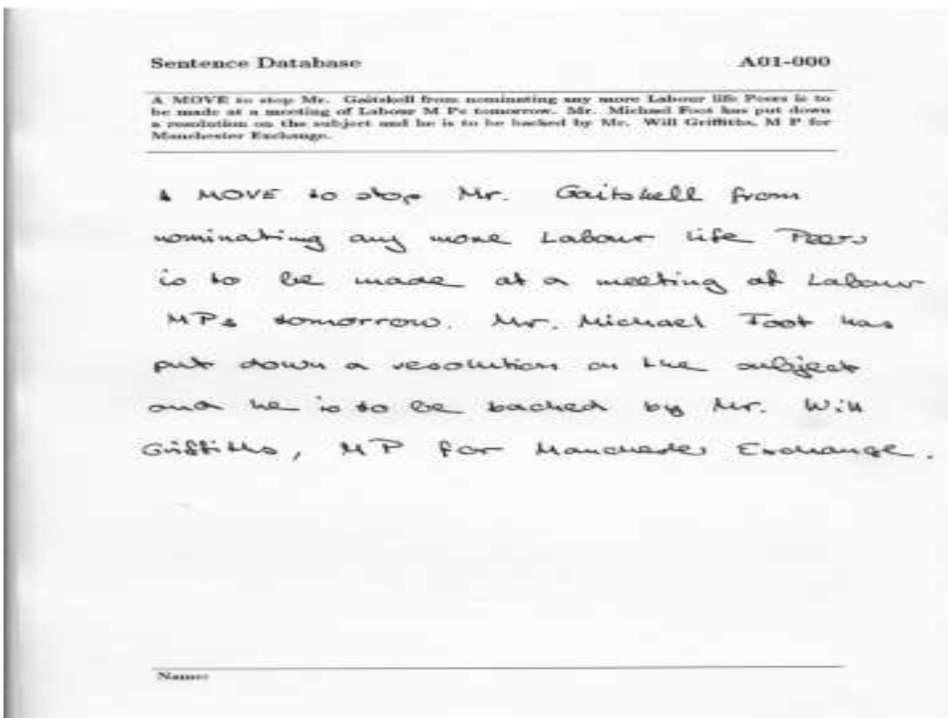
2.1 MOTIVATION

Handwriting is present in our daily lives, usually in notes, lists or other short texts in everyday life. But it is also used more systematically in other areas, such as taking notes in academic classes or in business meetings. Moreover, despite the advent of new technologies such as computers, tablets or smartphones, handwriting is still the preferred method for many people to capture their ideas or thoughts, at least initially. Handwriting can also be done virtually anytime and anywhere with a minimum of technology: with a pencil and a notebook, chalk on a wall or the hand in the sand.

2.2 TO THE DIGITAL AGE

In fact, the main disadvantage of the handwriting text is that it must be digitized to facilitate its preservation, arrangement, and dissemination. Now, we are in the digitization era. All the information is stored and indexed in digital formats, and all the business processes must be digital. With all data and knowledge stored in digital databases, multiple advantages are achieved in terms of accessibility and analysis of the information. In this context, the capacity to recognize and digitize the content of handwriting text is necessary to extract shareable knowledge from it.

3. DATA



Sentence Database A01-000

A MOVE to stop Mr. Gaitskell from nominating any more Labour life Peers is to be made at a meeting of Labour M Ps tomorrow. Mr. Michael Foot has put down a resolution on the subject and he is to be backed by Mr. Will Griffiths, M P for Manchester Exchange.

A MOVE to stop Mr. Gaitskell from
nominating any more Labour life Peers
is to be made at a meeting of Labour
MPs tomorrow. Mr. Michael Foot has
put down a resolution on the subject
and he is to be backed by Mr. Will
Griffiths, MP for Manchester Exchange.

Name:

An example forms from the IAM Handwriting dataset. Word images in the dataset were extracted from such forms.

Our main resource for training our handwriting recognizer was the IAM Handwriting Dataset. This dataset contains handwritten text of over 1500 forms, where a form is a paper with lines of texts, from over 600 writers, contributing to 5500+ sentences and 11500+ words. The words were then segmented and manually verified; all associated form label metadata is provided in associated XML files. The source text was based on the Lancaster-Oslo/Bergen (LOB) corpus, which contains texts of full English sentences with a total of over 1 million words. The database also includes 1,066 forms produced by approximately 400 different writers. This database given its breadth, depth, and quality tends to serve as the basis for many handwritings recognition tasks and for those reasons motivated our choice of the IAM Handwriting Dataset as the source of our training, validation, and test data for our models. Last but not least, in deep learning large datasets–even with many pre-trained models–are very important and this dataset containing over 100K+ word instances met those requirements (deep learning model need at least $10^5 - 10^6$ training examples in order to be in position to perform well, notwithstanding transfer learning).

3.1 DATA NORMALIZATION BACKGROUND

The data normalizations carried out on handwritten images are the following ones:

- Removal of background noise to improve the image contrast.
- Text slope correction.
- Text slant or cursive correction.
- Normalization in size of the ascenders and descenders zones in characters.
- Image resizing keeping its aspect ratio.

The main difficulty of the HTR problem lies in the large variability of handwritten text. Each individual has his/her own particularities when handwriting. Even the same person writes differently on different situations. This is why a great deal of effort has long been devoted to the development of text normalization algorithms to reduce this variability.

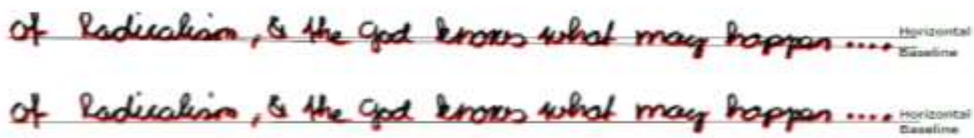
4. METHODS

4.1 IMAGE PROCESSING

Handwritten text exhibits a very high variability, which not only depends on the writer. Additionally, when the text is digitized, the quality and age of the paper and the digitization process itself can add significant noise to the image. That is why it is usual to pre-process the images of handwritten text before using them in models and recognition algorithms, which allows to eliminate noise and normalize the images to be analyzed as much as possible**.**

4.2 SLOPE CORRECTION

The slope or also called skew is the inclination of the lines of text with respect to a completely horizontal baseline. It usually appears when the text is written on a blank page without a pre-printed pattern of lines or boxes. The deviation of the line can have a positive angle if the writer tends to raise the line when he/she writes it, or it can also have a negative angle with respect to the horizontal if the line tends to descend. Slope correction can be performed at the page level, as well as at the line level or even at the word level. Likewise, the slope correction is sometimes carried out simultaneously with the slant correction since they are related, because a slope angle correction modifies the slant angle of the words by the same amount.



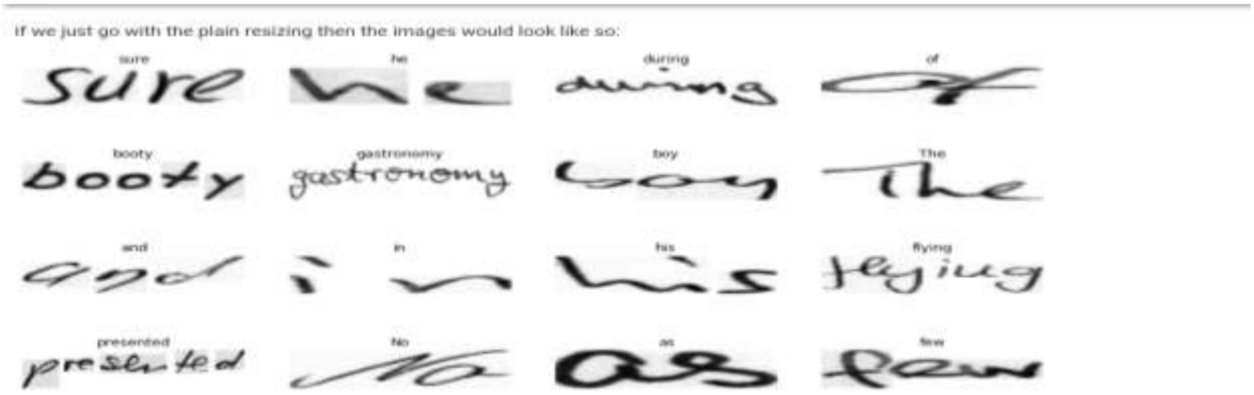
Example of slope angle estimation

4.3 SLANT CORRECTION

The slant identification and correction is a critical aspect in HTR since many recognition algorithms that use handwritten text images as inputs usually have an approach that manages the image column by column. For example, those that use dense neural network type models such as multilayer perceptron or recurrent neural networks that handle the image as a sequence of columns. Models using convolutional networks do not have this problem.

4.4 RESIZING IMAGES

Handwritten text recognition models based on direct image modeling (i.e., segmentation free), which are practically all used in recent years, usually need that input images to be fixed size, specifically fixed height. For this reason, a series of transformations are applied to a previously standardized line or word images to fit them into this size. These transformations usually include: centering the text in the image, adding borders around the text and resizing the image, keeping the aspect ratio (for which the image is usually completed with columns of white pixels on the right). The centering of the text in the image is done by removing the empty borders that do not contain any image strokes. Sometimes, a frame of blank rows and columns of fixed size is added around the text image to be centered. The image is then rescaled to a fixed height.



Resizing of images without distortion

4.5 SEGMENTATION

After the input images are pre-processed, individual characters are separated using a segmentation technique. These characters are then stored into a sequence of images. Then boarders in each character image are eliminated if the boarder is available. Segmentation of hand written text document into individual character or digit is an important phase in document analysis, character recognition and many other areas. Character segmentation has become a crucial step for mail address recognition in the automatic post mail sorting system.

Segmentation of the image

5. MODEL

5.1 CNN ARCHITECTURE

Convolutional Neural Networks (CNN) are a type of neural network specialized in modelling the information with a grid-like structure, such as images considered a two-dimensional pixel grid. These networks use the convolutional mathematical operation (*) instead of the standard matrix multiplication to calculate the network output.

Feature extraction is made on the segmented characters. In our case, the features are extracted using CNN with ReLU activation function. CNN works on each character image to form a matrix of reduced size using convolution and pooling.

- Here different convolutional layers, pooling layers, and dense layers are used.
  - The first one is LeNet. It was the first proposed for handwritten digit recognition in 1995. It is characterized by using 55 kernels, and having two sets of convolution and pooling layers, with an incremental number of feature maps.
  - The second one is VGG, which is characterized by proposing a convolution convolution-pooling layers with 33 kernels, and by duplicating the number of feature maps in each module.

– The third one is the ResNet architecture, which is characterized by proposing direct connections (residuals) between the input and output of each convolutional component. Additionally, it employs the regularization strategy called batch normalization.

5.2 CTC LAYER

Connectionist temporal classification (CTC) is a type of neural network output and associated scoring function, for training recurrent neural networks (RNNs) such as LSTM networks to tackle sequence problems where the timing is variable. They output a character-score for each time-step, which is represented by a matrix. We now need to use this matrix for:

- Training the Neural Network, i.e., calculating the loss
- Decoding the output of the Neural Network

Our model will use the CTC loss as an endpoint layer. The CTC algorithm is alignment-free. The objective function marginalizes over all alignments. While CTC does make strong assumptions about the form of alignments between XX and YY, the model is agnostic as to how probability is distributed amongst them. In some problems CTC ends up allocating most of the probability to a single alignment. However, this isn't guaranteed.

5.2.1 CTC WORKING

CTC works on the following three major concepts:

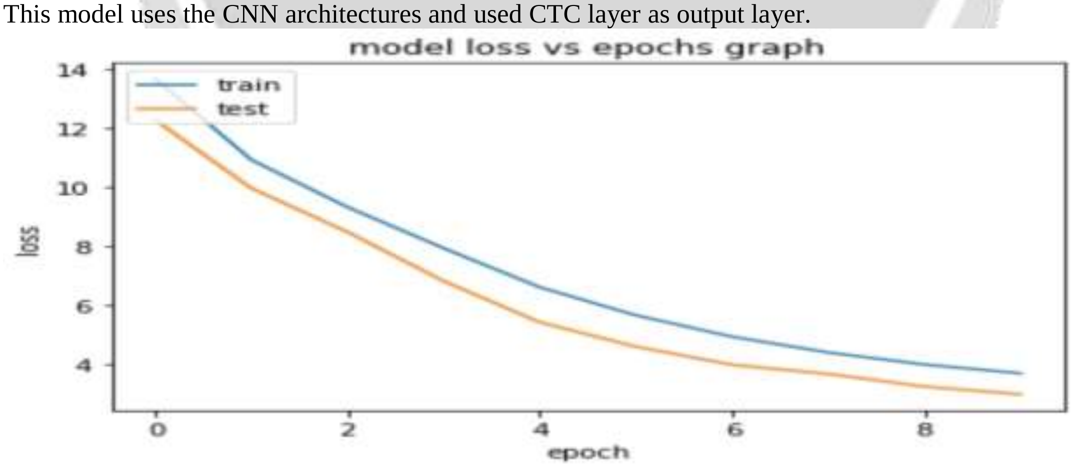
- Encoding the text
- Loss calculation
- Decoding

5.3 TRAINING AND RECOGNITION

The use of the same training, validation, and evaluation or test partitions of the reference datasets in the different investigations on any modeling problem is a fundamental aspect where there is a general consensus. The training partition is used in the parameter setting of the different experiments, the validation partition allows to compare the different experiments in terms of accuracy or error metrics to select the optimal parameterization. The evaluation or test partition allows to calculate the metrics values on the selected optimal parameterization experiment.

5.4 RESULTS

This model uses the CNN architectures and used CTC layer as output layer.



After making revisions to our learning rate and model input, we found that we found we had a consistently decreasing loss.

One technique we have found to be useful during experimentation was keeping the number of epochs low when trying out different hyperparameters. Since our training and validation processes took a long time, and we wanted to find the optimal hyperparameters, we trained our model on 10 epochs, compared the results, and reran our model on more epochs with a subset of hyperparameters that achieved the best results and the ones whose loss/accuracy graphs looked the most promising.

6. SCOPE OF HCR

Hand Written character recognition is one of the important fields of pattern recognition. within the scope of this area of important documents and archives and other written texts transferring to digital media. There are numerous applications of handwritten character recognition: reading postal addresses, bank check accounts and forms.

Handwritten character recognition is used in many cases like:

- **Electronic form filling:** One of the applications of online handwriting recognition is electronic form filling. Internationally, the expenditure for entry of data from handwritten forms, notes and records is trillions of dollars. All government application forms can be completed and filled using handwriting recognition and the data will be directly entered to structured databases.

- **Alternative to Hardware and Software Keyboards:** All keyboards have finite number of keys, so it limits the number of symbols that can be entered from the keyboard. But handwriting avoids such limitations and permits input of symbols (without limits), sketches and drawings. It facilitates writers to think freely while inputting the data.

- **Digitalization of palm leaf manuscripts:** As handwriting recognition deals with understanding the handwritten text which is already written on papers or other writable surfaces or materials. So one of problem is not solved yet. It is not even solved for Latin script completely.

- **Automatic conversion of prescription to typed form:** It is one of the challenging tasks to understand the doctor’s handwriting. This problem can be solved by using handwritten recognition. In case of automatic conversion of prescription to typed form, online handwriting will be employed.

7. CONCLUSION

- We found that the strategies of applying the proposed model directly on the image of each word or applying it on a sequence of patches extracted from each image provide results that favor the chunk partitioning strategy as the length of the image text increases.
- We found that normalizing the handwritten text images improves the accuracy of the proposed recognition model. Numerous tables with partial results that allow us to evaluate the individual contribution of each of the proposed improvements in the results of the new model have also been provided.

8. FUTURE WORK

- Preprocessing work is done in which normalization, filtration is performed using processing steps which produce noise free and clean output. Managing our evolution algorithm with proper training, evaluation other step wise process will lead to successful output of system with better efficiency. Use of some statistical features and geometric features through neural network will provided better recognition result of English characters. This work will be helpful to the researchers for the work towards other script.
- Experiment with other more recent network architectures that give good results on other Computer Vision problems, such as the EfficientNet models.
- Carry out cross-validations between the databases to better understand their generalization capacity. For this purpose, it should be taken into account that the different databases in origin have essential differences in terms of image size, border size, and whether they are binarized or not.

9. REFERENCES

[1]. http://cs231n.stanford.edu/reports/2017/pdfs/810.pdf
[2]. https://dl.acm.org/doi/10.1145/3395027.3419603
[3]. M. Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016. arXiv: 1603.04467 (cited on page 115).
[4]. T. Bluche et al. \Framewise and CTC training of Neural Networks for handwriting recognition". In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. Volume 2015-November. IEEE. 2015, pages 81{85. isbn: 9781479918058. doi: 10.1109/ICDAR.2015.7333730 (cited on pages 66, 67).