International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

Offline HandWritten Text Recognition

K. Sunitha[1], Ch. Vivek[2], Ch. Sowmya[3], V. Rajesh[4]
CSE Dept. , P V P Siddhartha Institute of Technology

---***---

Abstract - Handwriting recognition is one of area pattern recognition. The purpose of pattern recognition is to categorize or classify the data or object of one of the classes or categories. Handwriting recognition is defined as the task of transforming a language represented in its spatial form of graphical marks into its symbolic representation. Each script has a set of cons, which are known as or letters, which have certain basic shapes. The goal of handwriting is to identify input characters or image correctly then analysed to many automated process systems. The automatic recognition of handwritten text can be extremely useful in many applications where it is necessary to process large volumes of handwritten data, such as recognition of addresses and postcodes on envelopes, interpretation of amounts on bank checks, document analysis, and verification of analysis, and is needed to be able to read document or data for ease of document processing. Since the quality of clear reflected material images is much lower than printed or handwritten texts images that employ commonly used character segmentation and recognition algorithms, it is complicated to gain the high success rates by directly applying them to these images.

Key words: Classification, Recognition, Analise, Process.

1. Introduction

OFFLINE handwriting recognition is the task of determining what letters or words are present in a digital image of handwritten text. It is of significant benefit to man-machine communication and can assist in the automatic processing of handwritten documents [1]. It is a subtask of Optical Character Recognition (OCR), whose domain can be machine-print or handwriting but is more commonly machine-print [2]. The recognition of English handwriting presents unique challenges and benefits and has been approached more recently than the recognition of text in other scripts. This paper describes the state of the art of this field. Handwritten recognition is usually classified into two groups which are online and offline. Online character recognition deals with information about writing dynamics as the text is being written while offline character recognition deals with static Information in which acquisition is done after all the text is written Offline character recognition usually uses other medium of written text such as papers. One of the main issues in handwritten text recognition is that its accuracy, in human depends on knowledge about which language the text is written [3]. Same text of the same corpus but written in different language can result in different accuracy. It is "offline" if it is applied to previously written text, such as any images scanned in by a scanner. The online problem is usually easier than the offline problem since more information is available [4].

2. System Analysis:

2.1. Existing System

Since a very long time, human used to write their thoughts in the form of letter, transcripts etc.; to convey them to others. But since the development of computer technology the format of handwritten text changed rapidly to computer generated to develop such system in past. Though, there is still a need of much more research in this field. Many recognition studies have been made for offline and online handwritten characters of major languages used worldwide: like English, Chinese, Indian scripts such as Devanagari, Malayalam, and Bangla 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. [6]. Thus, recognition studies of handwritten character image samples remain relevant because of their enormous application potentials.

2.2. Proposed System

We use a NN for our task. It consists of convolutional NN (CNN) layers, recurrent NN (RNN) layers and a final Connectionist Temporal Classification (CTC) layer. Handwritten text classifiers were first required for classification of postal mail. Using scanning equipment, hardwired logic recognized mono-spaced fonts [7]. The first Optical Character Recognition (OCR) software developed in 1974 by Ray Kurzweil. By reducing the problem domain, the process was more accurate. This allowed for recognition in handwritten forms. Foremost, it lacked efficiency and knowledge of unexpected characters.

2.3. Proposed System

We use a NN for our task. It consists of convolutional NN (CNN) layers, recurrent NN (RNN) layers and a final Connectionist Temporal Classification (CTC) layer. Convolutional neural networks (CNN) are all the rage in the deep learning community and domains, and they're especially prevalent in image and video processing projects. RNNs share the parameters across different time steps [8]. This is popularly known as Parameter Sharing. This results in fewer parameters to train and decreases the computational cost. CNN also follows the concept of parameter sharing. A single filter is applied across different parts of an input to produce a feature map as shown in Fig -2.1.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

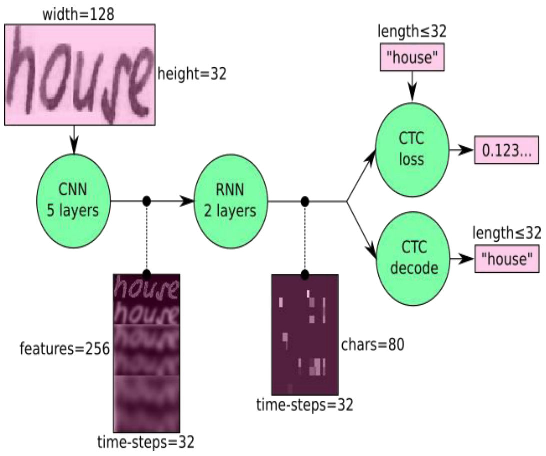

Fig -2.1: Overview of the NN operations (green) and the data flow through the NN (pink)

We can also view the NN in a more formal way as a function (see Eq. 1) which maps an image (or matrix) M of size W×H to a character sequence (c1, c2, …) with a length between 0 and L. As you can see, the text is recognized on character-level, therefore words or texts not contained in the training data can be recognized too (as long as the individual characters get correctly classified). Recurrent Neural Networks or RNN as they are called in short, are a very important variant of neural networks heavily used in Natural Language Processing. In a general neural network, an input is processed through a number of layers and an output is produced, with an assumption that two successive inputs are independent of each other as shown in the Fig -2.2.

$$\text{NN}: \underset{W \times H}{M} \rightarrow \underset{0 \le n \le L}{(c_1, c_2, \ldots, c_n)}$$

Fig -2.2: The NN written as a mathematical function which maps an image M to a character

Operations:

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 can 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. The IAM dataset consists of 79 different characters, further one additional character is needed for the CTC operation (CTC blank label), therefore there are 80 entries for each of the 32 time-steps.

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.

3. UML Diagrams

3.1. Use Case Diagram

The basic difference between a feed forward neuron and a recurrent neuron is shown in figure 1. The feed forward neuron has only connections from his input to his output. In the example of figure 1 the neuron has two weights. The recurrent neuron instead has also a connection from his output again to his input and therefore it has in these example three weights. This third extra connection is called feed-back connection and with that the activation can flow round in a loop. When many feed forward and recurrent neurons are connected, they form a recurrent neural network shown in the following Fig -2.3.

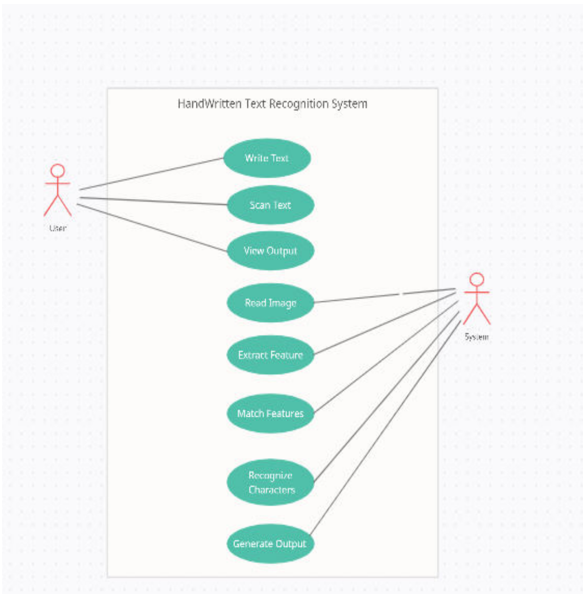

Fig -3.1: Basic UML diagram of the text recognition process

3.2. Class Diagram

This architecture has also been used on speech recognition and natural language processing problems as shown in Fig -2.4 where CNNs are used as feature extractors for the LSTMs on audio and textual input data. It continues to be a challenging problem for an automatic recognition system to detect text from these images because they may face unexpected conditions such as closing loops, spurious branches, and shiny or raised text. Other conditions could be

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

related to the surface on which the text is printed, such as curvature, indentation, or matte finishes.

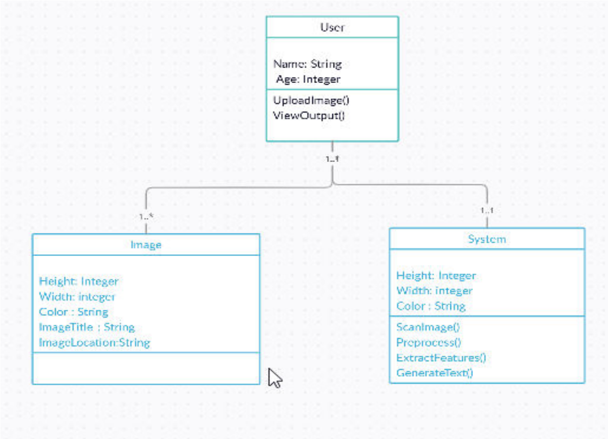

Fig -3.2: Class diagram of the text recognition process

3.3. Sequence Diagram

CNNs employ filters within convolutional layers to transform data. Whereas, RNNs reuse activation functions from other data points in the sequence to generate the next output in a series. While it is a frequently asked question, once you look at the structure of both neural networks and understand what they are used for, the difference between CNN and RNN will become clear.

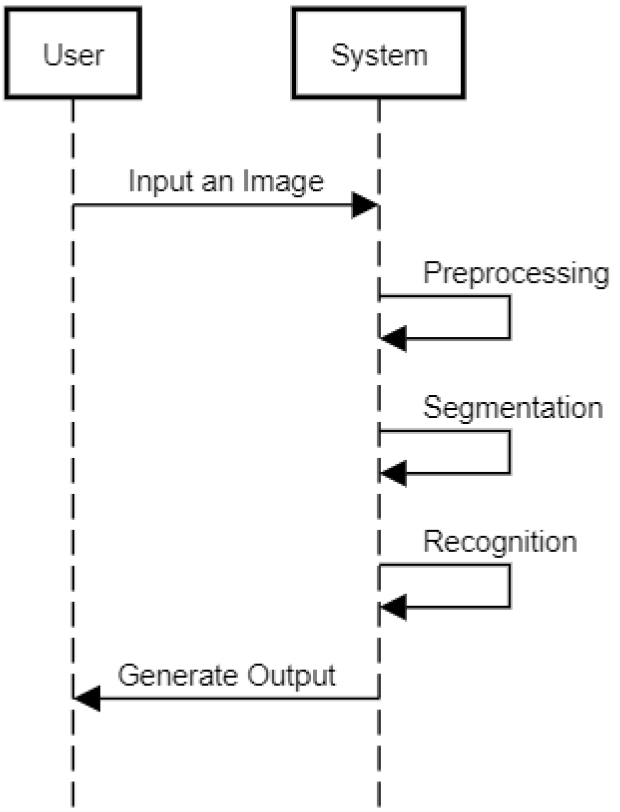

Fig -3.3: Sequence diagram of the text recognition process

3.4. Activity Diagram

The recurrent connections often offer advantages. They make every unit to use their context information and especially in image recognition tasks this is very helpful. As the time steps increase, the unit gets influenced by larger and larger neighborhood. With that information recurrent networks can watch large regions in the input space. In CNN this ability is limited to units in higher layers. Furthermore the recurrent connections increase the network depth while they keep the number of parameters low by weight sharing as in Fig -2.6.

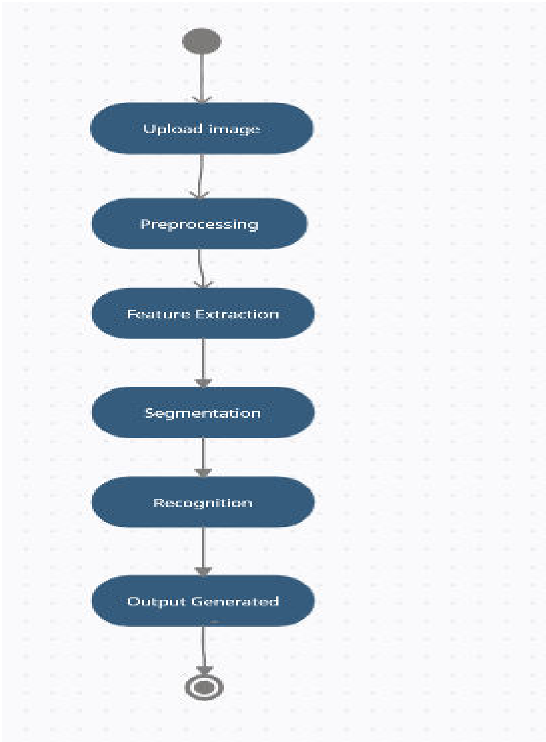

Fig -3.4: Sequence diagram of the text recognition process

4. Results

A text recognition system learns whatever is helpful to increase the accuracy in the dataset it is trained on. If some random looking pixels help to identify the correct class, then the system will use them. And if the system only has to handle left-aligned text, then it will not learn any other type of alignment. Sometimes, it learns features which also we humans find useful for reading and which generalize to a wide range of text styles, but sometimes it learns short-cuts which are only useful for one specific dataset. The results are shown in Fig -4.1.

Fig -4.1: Test Image

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

Another interesting property of the score function is the periodicity of cursive writing as shown in Fig -4.2. These four pixels equal the downsizing factor of the convolutional network from a width of 128 pixels to a sequence length of 32.

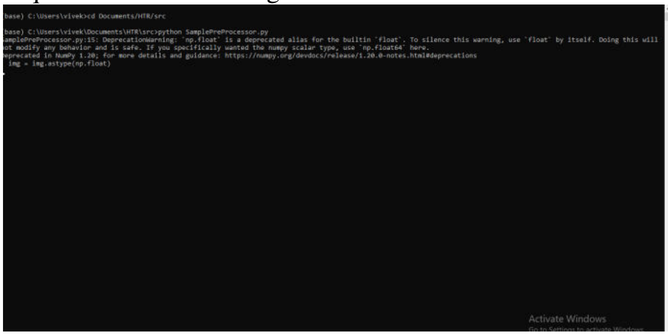

Fig -4.2: Pre-Processing the test image

Further investigations are needed to explain this behavior, but it might be caused by the pooling layers with their discontinuities: shifting a pixel one position to the right, it might stay within the same pooling cluster, or it might step over to the next one, depending on its position.

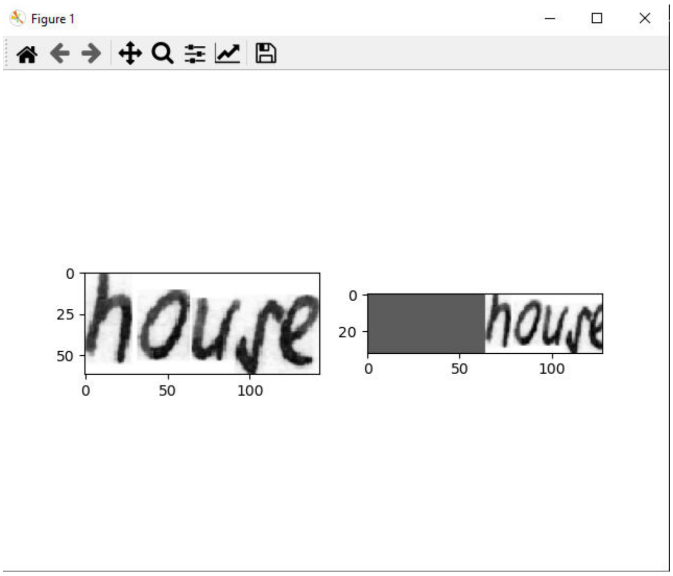

Fig-4.3: Transforming the Test Image

Fig -4.2 shows samples of enhancement results.

The last column in the figure shows the best result (in terms of image quality and intensity difference between text and background). In comparison, the new methods illustrate that conventional enhancement (histogram, linearization, and local deviations) does not yield successful results even for fairly clean images, as shown in Fig -4.4.

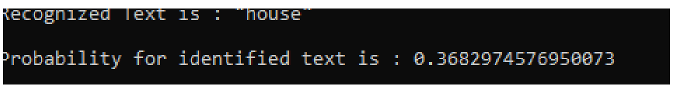

Fig-4.4: Output of the model

5. Future Scope

In this project classification of characters takes place. The project is achieved through the conventional neural network. This algorithm will provide both the efficiency and effective result for the recognition. The project gives best accuracy for the text which has less noise. The accuracy completely depending on the dataset if we increase the data, we can get more accuracy. If we try to avoid cursive writing, then also its best results.

6. Conclusion

We discussed a NN which can recognize text in images. The NN consists of 5 CNN and 2 RNN layers and outputs a character-probability matrix. This matrix is either used for CTC loss calculation or for CTC decoding. An implementation using TF is provided and some important parts of the code were presented. Finally, hints to improve the recognition accuracy were given.

References

1. Marti, U-V., and Horst Bunke. "The IAM-database: an English sentence database for offline handwriting recognition." International Journal on Document Analysis and Recognition 5.1 (2002): 39-46.
2. Mithe, Ravina, Supriya Indalkar, and Nilam Divekar. "Optical character recognition." International journal of recent technology and engineering (IJRTE) 2.1 (2013): 72-75.
3. Ingle, R. Reeve, et al. "A scalable handwritten text recognition system." 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019.
4. Kim, Gyeonghwan, Venu Govindaraju, and Sargur N. Srihari. "An architecture for handwritten text recognition systems." International Journal on Document Analysis and Recognition 2.1 (1999): 37-44.
5. Parvez, Mohammad Tanvir, and Sabri A. Mahmoud. "Offline Arabic handwritten text recognition: a survey." ACM Computing Surveys (CSUR) 45.2 (2013): 1-35.
6. Hanifa, Rafizah Mohd, Khalid Isa, and Shamsul Mohamad. "A review on speaker recognition: Technology and challenges." Computers & Electrical Engineering 90 (2021): 107005.
7. Abdulrazzaq, Maiwan Bahjat, and Jwan Najeeb Saeed. "A comparison of three classification algorithms for handwritten digit recognition." 2019 International Conference on Advanced Science and Engineering (ICOASE). IEEE, 2019.
8. Salazar, Julian, Katrin Kirchhoff, and Zhiheng Huang. "Self-attention networks for connectionist temporal classification in speech recognition." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.