International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

A Review of Literature on Handwritten Text Recognition

Swati Sinha[1], Yash Gurav[2], Priyanka Bhagat[3], Rajeshri Jadhav[4]
[1]*Professor, Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India*
[2,3,4]*Student, Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India*

***Abstract*: Optical Character Recognition is used to recognize text in the documents and convert it into an editable machine-encoded format. An OCR system depends mainly on feature extraction, recognition and classification into appropriate labels. This paper describes segmentation, recognition and identification of handwritten documents using image processing and CRNN based architecture. Further, a decoding algorithm, Word Beam search (WBS) is presented that helps decode the features into label sequences with the highest degree of probability.**

***Keywords*: OCR, ROI, CNN, RNN, Noise.**

1. Introduction

Character recognition is an art of detecting, segmenting and identifying characters from the image. Character recognition is mainly of two types online and offline. In online character recognition, data is captured in real-time using a device. In offline recognition, a prewritten document is used for data capturing. [1]

An OCR system depends mainly upon the process of feature extraction and classification. Since every individual possesses a different written behaviour this system possesses a challenge that needs to be tackled by various technologies available today. Accurate working of such a dynamic system is still an area of research. In this decade researchers have integrated machine learning and deep learning approaches with image processing to appropriately classify the images. Various approaches like Support Vector Machine (SVM), Nearest Neighbour Classifier (NNC), Hidden Markov Model (HMM) and Neural Networks with its variations have been implemented.

2. Existing System

Several OCR systems have been developed and each of them varies based on the image processing, feature extraction, classification techniques used.

The basic image pre-processing steps are standard in all character recognition software.

The steps include:

1) Converting the RGB based image into greyscale.
2) Converting a grayscale image into a binarized format.
3) Identifying the characters and resizing them into the size by which the model is trained. A small amount of noise gets removed when converting the image from grayscale into the binarized format due to the threshold value set. For feature extraction and classification techniques, various methods have been proposed.

One of the methods proposed in the paper by SK Patel & J Jha [1] includes basic steps needed to be implemented in the OCR systems. Several machine learning techniques like nearest neighbours, support vector machines, HMMs and RNNs have been compared for accuracies. RNNs have emerged as one of the best methods for classification. One of the systems has used Artificial Neural Networks to recognize digits from the images.[2]

A further step has been taken in the paper by [3]. It compares the various neural network approaches used in the OCR systems. The result shows a deep convolutional neural network (DCNN) has better capability to extract features.

A CRNN based approach has been implemented in the paper presented by B. Shi, X. Bai and C. Yaowhich [4] uses LSTMs for sequential processing of the image and thus reduces possible out of context errors.

3. Critical Analysis of Existing System

- The existing systems have difficulty in differentiating between noise and a word. A small dot which is actually a noise also gets detected as a word.
- If text is on any ruled line paper then those lines will be considered as some potential words.
- The resized words are not again processed in any existing system. There is a loss of pixels while resizing and thus the characters in the word break and are not classified properly by the model, reducing the accuracy.
- The system uses CRNN architecture but does not uses proper decoding algorithms which causes linguistic and contextual errors.

4. Proposed System

A. Ruled line removal

The ruled lines can be detected using the Hough line transform presented in the paper by M. A. A. Refaey [5]. A binarized image is to be passed through this algorithm. The algorithm works by counting the number of pixels forming a

straight line, so for different images captured from different sources, it is difficult to set a single threshold value. A simple solution to this can be resizing the image to a standard size before starting any image processing on it. For the algorithm to effectively detect lines canny edge detection can be used.

B. Noise removal

A method presented in the paper by Z. D. Kun [6] can be used to remove the noise. In this, different filters are used for opening and closing operation both hence the filters can be set accordingly to get a noise-free binarized word. Also, based on analysis the noise will have a much smaller area compared to a word thus any segmented character can be treated as noise if its area is less than a threshold value.

C. Processing on segmented word

The region of interest can be identified using a region growing algorithm which can be done in python using findContours() method. Thus, the ROI can be resized to a pre-determined size. After resizing, performing dilation using appropriate filter thickens the word boundaries and fixes any small breaks in words. Thickening the word can effectively increase the efficiency of recognition.

D. CRNN architecture

The project uses CRNN model for recognizing English words. It is a LSTM based architecture [7] for recognizing sequential features from the input image. CNN has been used in various contexts including scene text recognition, object recognition, etc. Deep CNNs are trainable and output feature maps that can be used for the various visual recognition task.

RNNs are advantageous in terms of contextual understanding and for semantic correctness which won't be possible if we process one character at a time. It operates on sequences of arbitrary length from left to right.

The images are pre-processed and scaled to the same height. It is then inputted into the CNN architecture. The CNN architecture includes convolutional layers, pooling layer and uses RELU activation function which outputs a feature sequence of size 32x256.

RNN architecture takes feature sequence as input, each with 32-time stamps. The feature sequence contains 256 features per time-steps, the RNN propagates relevant information through this sequence. Each column of the feature vector corresponds to the rectangular region of the original image.

An RNN outputs predictions for each of the 32 frames which is then converted into a probability for the 80 label sequences. Label sequence with the highest probability is outputted. The system is trained with the CTC loss function which enables training of data and targets

E. Decoding algorithm

The decoding of the predictions in the corresponding label sequence is done using decoding algorithms like Vanilla beam search, Word beam Search. As proposed by Harald Scheidl in his paper [7], word beam search is more accurate than vanilla beam search and it constraints words to be used from the dictionary, thereby reducing the error rate. WBS uses the prefix tree to further output the correct labels. The system is thus developed using the CRNN architecture and the Word Beam Search Decoding algorithm.

5. Conclusion

A number of OCR systems were studied and accurate methods, techniques were analyzed. OCR is a challenging area of research and introduces new challenges with its changing dynamics. Accuracy and proper contextual conversion need to be checked for satisfying successful implementation of an OCR system.

Acknowledgement

We express our profound gratitude and sincere thanks to Dr. Swati Sinha for her pivotal support and guidance in the project and this paper.

References

[1] S. R. Patel, and J. Jha, "Handwritten Character Recognition using Machine Learning Approach - A Survey" International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), 2015.
[2] K. T. Islam, G. Mujtaba, R.G. Raj, H. F. Nweke, "Handwritten Digits Recognition with Artificial Neural Network" International Conference on Engineering Technology and Technopreneurship (ICE2T), 2017.
[3] M. M. J. Ghosh, A. Y. Maghari "A comparative study on handwritten digit recognition using neural networks," International Conference on Promising Electronic Technologies, 2017.
[4] B. Shi, X Bai and C Yaowhich, "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition," 2015
[5] Mohammed A. A. Refaey, "Ruled Lines Detection and Removal in Grey Level Handwritten Image Documents",6th International Conference on Information and Communication Systems (ICICS), 2015.
[6] Zhang Da-kun, "An Extended Opening Operation and Its Application in Image Processing", International Conference on Multimedia and Information Technology, 2008.
[7] H. Scheidl, S. Fiel, R. Sablatnig, "Word Beam Search: A Connectionist Temporal Classification Decoding Algorithm",16th International Conference on Frontiers in Handwriting Recognition, 2018.