International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 1737–1743

Survey on Devanagari Text Detection in Deep Learning

Aditya Sovani, Ashish Pandita, Chinmay Bansod, Vina Lomte [4]

[1]Department of Computer Engineering , RMD Sinhgad School of Engineering.
[4]Head of Department, Department of Computer Engineering , RMD Sinhgad School of Engineering.

Abstract

Since past decade, numerous attempts are made to digitize Indian languages and scripts. Various computer research organizations are working on digitization of Indian languages. There are Devanagari datasets being developed to recognize text or content in the input. A survey of the Devanagari word recognition system online handwritten is discussed in depth in this article. The increased use of handheld devices that accept handwritten data as input created a need for an application that effectively analyzes and acknowledges data. A camera is used as an input device, owing to the popularity of the digital interface. Using the feature CNN acknowledges the word or letter. This survey focuses on various technologies available on detecting Devanagari text.

Keywords: *Machine Learning, Deep Learning, CNN, RCNN, Character Recognition, Gesture Detection.*

1 Introduction

The process of converting image content to machine readable or computer encoded text is known as Optical Character Recognition(OCR). OCR is relatively older technology existing since mid 70s and early 80s. From the early 2000s, deep neural networks have been used with significant success in the field of image processing for object detection, segmentation, and recognition. Recognizing handwritten Devanagari characters has always been a difficult task. There are various tools and methodologies to recognize handwritten text such as Fuzzy based classification [2], Support Vector Machine [1], Markov Model [3].

The main benefit of the deep neural network is that, with the assistance of the brain, it needs less engineering. Use different algorithms it discovers the capabilities of the input data available on its own. Through data development the learning earns higher and stronger. Additionally it includes knowledge variations. When the samples are different in numbers, the one variants must be checked, otherwise the new samples will be misclassified.

ISSN: 2233-7857 IJFGCN

There are usually three ways to recognize Devanagari text: From images, via handwriting on a digital screen and through gestures in front of webcam or similar device.

1.1 Script Representation

Devanagari is among the scripts that is the basis for almost languages & 120 Indo-spoken to be exact spoken in South-East Asia. Devanagari mainly includes well known languages like Sanskrit, Hindi, Marathi, Maithili, Awadhi, Newari & Bhojpuri. It is used in writing classical Sanskrit text. All the Indian ancient mythological books are written in Devanagri script. This is made up of 10 digits and 47 alphabets, 14 of those 47 are vowels and 33 are consonants. Table 1 below displays all Hindi-language consonants in their base form. The following table 2 indicates all the vowels in the Marathi language, and the numerals used in the Devanagari script are shown in table 3. [10]

Table. 1 Basic Form of characters[11]

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ
ड	ढ	ण	त	थ	द	ध	न	प	फ	ब	भ
म	य	र	ल	व	स	ष	श	ह	क्ष	त्र	ज्ञ

Table. 2 Devanagri Vowels[11]

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	अं	अः

Table. 3 Devanagri Numerals[11]

०	१	२	३	४	५	६	७	८	९

ISSN: 2233-7857 IJFGCN

2 Literature Review

Most of the related work to the topic makes use of deep neural networks and deep learning.

A Convolutional Neural Network transforms the learnt features with data input and uses layers of 2D-convolution for producing output. Finally the objects are detected from those structures[4].Though deep learning can be used to work on problems encountered by AI community, it has massive success in the area of Natural Language Processing [5].

DHCD (Devanagari Handwritten Character Dataset.) is a public image dataset contraining over 92,000 consisting of 56 different classes. CNN (Convolution Neural Network) is subclass of DNN. Input can be correctly modeled by changing number of hidden layers and trainable parameters [6].

Research teams have been using various methods to address handwriting rec-ognition issues, with some attempting to combine different recognition algorithms to resolve other algorithm's weak areas. Still, as one algo-rithm, neural network is powerful enough to recognize handwriting samples and achieve high rate of recognition. NN(Neural Network) has capabilities for self-learning, self-adaptation and self-processing, making it reliable for scripted systems for character recognition [7]. Within the typical neural network there are three layers, which are the information layer, the output layer and thus the hidden layer. Both layers refer to coordinate the preparation of sets for the measurement of acknowledgments. The Neural Networks are also termed as perceptron multilayers. One benefit of using neural networks is that they are mostly equipped to follow the law of learning error-correction [8].

2.1 Work done on Devanagari Text Detection from Images.

Image based text can be detected by segmenting and recognizing a character. There are four phases in character segmentation. Recognition has testing and training phase [11]. This project involves, Input image, pre-processing, binarization, boundary detection, segmentation etc.

2.2 Work done on Devanagari Text Detection via handwriting.

Traditional neural network algorithms are used to process multiple formed array data. In this work 3 layers of CNN are used. Devanagari handwriting can also be classified via fuzzy based classification [2]. That system uses fuzzy classification on pre-processed characters before extracting features. Deep learning enables mathematical models consisting of multiple layers of processing to learn information interpretations with various abstraction levels. Such techniques have improved significantly the latest technology in voice recognition, optical image recognition, object recognition and several other fields[3].

ISSN: 2233-7857 IJFGCN

2.3 Work done on Text Detection via gestures through webcam.

There isn't much research available on such domain, but there might be some notable work which might be implemented but not formally documented. There are some projects to detect hand gestures via webcam. There is one project to detect computer commands via gestures. This project has different operations assigned to a gesture. For example, click for one finger up, cursor to right for two fingers. Cursor down for three fingers etc. OpenCV is used to detect Convexity Defects [14

3 Methodologies

3.1 System Architecture

Recognition of the letter / gesture character typically starts with the picture of the letter that the scanner or camera obtains. Data objects obtained from these methods must made ready for processing before they can be used for detection. Preprocessing involves steps such as the reduction of noise, duplication and distortion detection. This starts with repetition, where the input is a binary representation of gray or colour.

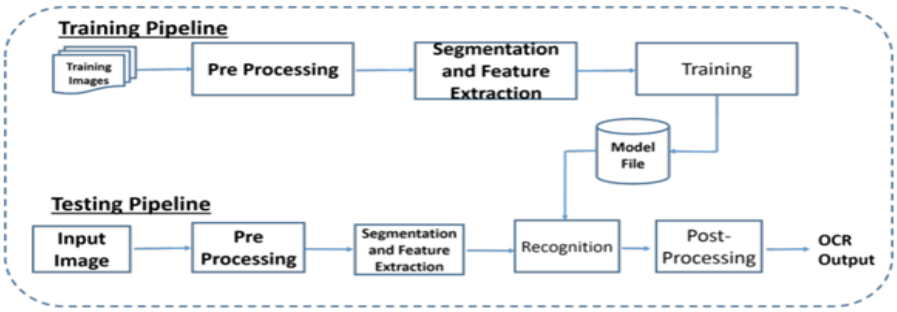

Fig. 1 Training & Testing flow.

Throughout Figure 1, an OCR typically uses a post processing unit to improve the classifier's classification errors. The post-processing unit makes use of models of other languages as well as dictionary very much for that reason.

ISSN: 2233-7857 IJFGCN

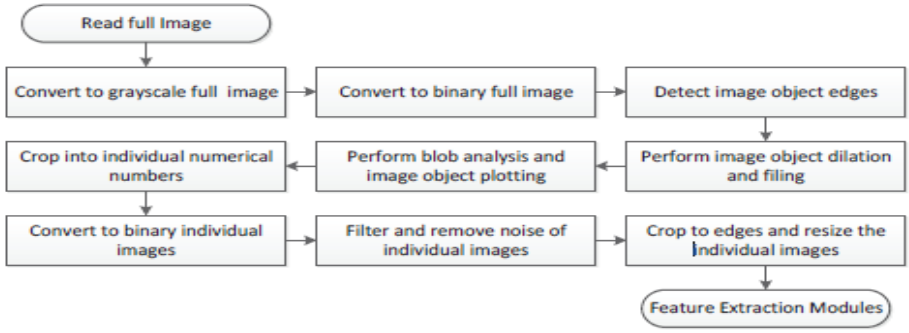

Fig. 2 Raw Processing of Image before actual processing[11]

Once a picture is put into the handwriting recognition device, even from a web device, it is important that the image is processed using standardized image processing methods for quick and accurate data capture.
Steps followed by the System:

The first phase is preprocessing. This phase includes changing the image data in a certain way as to be acceptable for the next phase.

The second phase is segmentation. This phase separates every individual letter within the image that will then be resampled to a scale specified by the instance of the learning.

The third phase is feature extraction. The differentiated picture will be used to remove the character's characteristics.

There are other methods of segmentation, extraction techniques that pick the easiest attributes of all such existing so on to achieve greater precision.

3.2 Recognition Controversies

The key problems for the Indian scripts when constructing a recognizer and classifier is that they need an outsized amount of symbols compared to languages such as English. English primarily comprises 52 characters while a script like Devanagari has about two hundred symbols.

Many of the Sanskrit languages are nearly similar to others. A number of them have the characteristic of having the combination of characters to construct another, commonly referred to ' jodakshara' in which Joda is combined/ Joint and 'akshara' is single alphabet. The detection of vowel modifiers or 'kanamatras' is another problem.

The Vowel modificators change the character's UNICODE meaning. Thus in order to detect the proper character and its UNICODE meaning, there are given some unique rules for the script. Such modifiers may also be there at the character's left, right, bottom or top. In a variety of instances, the

ISSN: 2233-7857 IJFGCN

modifier on an identical character may also be present at two positions. It is therefore important to correctly classify such modifiers in order to reduce the error in the classifiers.

For Indic languages, several of the optical character recognizers incorporate such that they disintegrate the whole word into equivalent symbols and are then recognised individually. It is done in accordance with the manner in which they appear, but when UNICODE is determined for these symbols, they must be rearranged and often combined to obtain UNICODE. The certain reason that after the consonants the' matra' or vowel modificator will appear is why we would like to rearrange the symbols. Figure 3 demonstrates the UNICODE reorganization process which is performed for proper rendering of the output. [10].

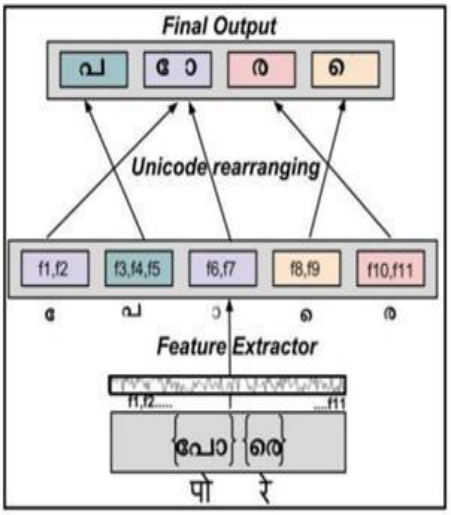

Fig . 3 Process Flow of UNICODE rearrangement [10]

4 Algorithmic Survey

4.3 CNN Family of Algorithms

Convolutional neural network (are widely used for CNNs) is one among the most cate-gories to try to to image recognition, image classifications, object detection, recognition faces etc., are a number of the areas where CNNs are widely used.

ISSN: 2233-7857 IJFGCN

Image classification using CNN is a popular application. Devices interpret an input picture as a pixel range and that depends on the size of the picture. Based on the image resolution, it will see h x w x d where h stands for Height of the image, w stands Width of the image, d stands for Dimension of the image.

We surveyed on a paper that has proposed Neural Networks sponsored an groundbreaking character classifier model to interpret the text regions obtained from the extraction step[15]. The genetic algorithm is trained on images of individual english words of 32 * 32 pixels in size. The collection includes photographs taken from signboards, flags, banners, drawings, and certain realistic images. Therefore it comprises very vibrant and free style-drawn texts for every character which is perfect for us[15]. There weren't enough pictures for each letter, so we cropped the pictures manually. The manually cropped images were used, because of less photos available for each letter. Chars74K dataset is also used apart from IIIT 5K-word dataset. It introduced a new technique for detecting and acknowledging text from scene pictures. The method performs very well and produces good performance with reasonable precision over a variety of images [15].

4.4 Regional Convolutional Neural Network (R-CNN)

We have tried out many algorithms to build the most precise model in the least amount of time. And this journey, spanning multiple papers and real-world datasets, it led us to the R-CNN family of algorithms. Apart from CNN based approaches, more precise model can be built in least amount of time by the R-CNN family of algorithms.

Text detection using gestures requires focusing on any particular point from the input image. All other parts are to be excluded except for the point from which the gesture is being drawn. In CNN, it requires large amount of data for its processing. Due to this it becomes a complex task to focus on single point in CNN. Whereas in R-CNN a particular region is selected from the data and further processing is done. Thus, using R-CNN it becomes more faster as well as accuracy of the output also increases. Thus R-CNN should be preferred over CNN algorithm.

CNN was quickly evolved into R-CNN and its further work to fast R-CNN and faster R-CNN. It utilizes a technique called Region of Interest Pooling allowed for sharing expensive computations and made the model much faster.

The Faster R-CNN also seems to be ideal for accurate object detection[16], fast & accurate start-to-end CNN. Consistently learnable, start-to-end text identification approach is suggested based on the Faster R-CNN[16]. The suggested training approach was explained as being more successful than the

ISSN: 2233-7857 IJFGCN

Faster R-CNN's traditional strategy[16]. The new training approach makes the traditional Faster R-CNN approach more effective.16]

Image text identification could be difficult, as the input may have various alignments, sizes, font types, low light, distortions of perception and languages.[17] It solves the issue by developing a Rotational Region CNN (R2CNN) [17]. In this paper, Rotational Region CNN (R2CNN) is proposed as a image text identification tool to address the challenges of randomly directed image text identification.[17] Based on Faster R-CNN, R2CNN is suggested for the identification of randomly focused image texts.[17]

5 Applications

Gesture Detection can be used to detect and recognize words drawn in motion. Also using gesture detection recognition of joint letters can also be done. Gestures can be used to interact with the UI of a Smart TV. High-end smart TV remotes already use motion tracking to help users use their remote like a pointer. Typing using remotes is often a tedious task and requires a lot more time, which by implementing Gestures can make our task a lot easier, fast & accurate. Interacting with Displays/Kiosks at public spaces would be a lot easier.

One of the important applications for Gesture Based Text Detection can be for the Blind People. Gesture Based Text Detection can enable the blind people for entry into the digital world all by themselves. Blind People would be enabled to digital writing. This could prove to be a revolution for the blind community.

6 Conclusion

An hand-written Devanagari word recognition program using CNN is surveyed in this survey paper. The main purpose of this paper is to study CNN algorithm in extraction of the Devanagari functionCNN can identify the word by using these tools. We also observed that techniques for identifying and detecting Devanagari characters from pictures, as well as handwriting are valid.

References

[1] Prasad K. Sonawane,Sushama Shelke, “Handwritten Devanagari Character Classificationusing Deep Learning”, Aug 2018.

[2] S. Shelke and S. Apte, “A Fuzzy based Classification Scheme for Unconstrained Handwritten Devanagari Character Recognition.”, Jan 2015

ISSN: 2233-7857 IJFGCN

[3] B. Shaw, S. Parui and M. ShridharB. Smith, “An approach to graphs of linear forms (Unpublished).”

[4] Y. LeCun, Y. Bengio and G. Hinton , “Deep Learning,” May 2015

[5] Collobert, R., et al., “Natural Language Processing (Almost) from-Scratch,” Journal of Machine Learning Research, 2011.

[6] “Deep Learning Based Large Scale Handwritten Devanagari Character Recognition.”, 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA),2015

[7] Y. Tai-Shan, T. Yong-Qing, and C. Du-wu, "Research on handwritten numeral recognition method based on improved genetic algorithm and neural network.",2014

[8] W. Xin, T.-l. Huang, and X.-y. Liu, "Handwritten Character Recognition Based on BP Neural Network,",2015

[9] Kai Ding, Zhibin Liu, Lianwen Jin, Xinghua Zhu, “A Comparative study of GABOR feature and gradient feature for handwritten chinese character recognition”, Nov 2007

[10] A. Ray, A. Chandawala and S. Chaudhury, “Character Recognition using Conditional Random Field Based Recognition Engine”, 2013

[11] Parul Sahare and Sanjay B. Dhok, “Multilingual Character Segmentation and Recognition Schemes for Indian Document Images”

[12] Jasrotia, D. K., & Malik, A. (2018). OCR, Webcam, Template Matching Feature

[13] Extraction. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 7(8), 216-222Jasrotia, D. K., & Malik, A. (2018). OCR, Webcam, Template Matching Feature Extraction. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 7(8), 216-222Jasrotia, D. K., & Malik, A. (2018). OCR, Webcam, Template Matching Feature Extraction. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 7(8), 216-222

[14] Athiya Marium, Deepthi Rao, Divina Riya Crasta, Kavya Acharya, Rio D’Souza, Department of Computer Science and Engineering, St Joseph Engineering College, Mangaluru, India,“Hand Gesture Recognition using Webcam”, 2017

[15] Savita Choudhary, Nikhil Kumar Singh, Sanjay Chichadwani, “Text Detection and Recognition from Scene Images using MSER and CNN”, 2018

ISSN: 2233-7857 IJFGCN

[16] Yoshito Nagaoka, Tomo Miyazaki, Yoshihiro Sugaya, Shinichiro Omachi, "Text Detection by Faster R-CNN with Multiple Region Proposal Networks", 2017

[17] Yingying Jiang, Xiangyu Zhu, Xiaobing Wang, Shuli Yang, Wei Li, Hua Wang, Pei Fu and Zhenbo Luo, "R^2CNN: Rotational Region CNN for Arbitrarily-Oriented Scene Text Detection"

ISSN: 2233-7857 IJFGCN