IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10713

Search Algorithm for Kannada Handwritten Characters

Ms. Nischitha S[1], Ms. Lekhana G & Ms. Namratha G[2], Ms. Keerthana M & Ms. Meghana M K[2]

Assistant Professor, Dept of Electronic and Communication engineering, BGS Institute of Technology, B G nagar, Mandya, Karnataka, India[1]

Dept of Electronic and Communication engineering, BGS Institute of Technology, B G nagar, Mandya, Karnataka, India[2]

ABSTRACT: Kannada handwritten reports were the only way of documentation available in government offices and healthcare departments in Karnataka state. Reproducing the contents of these old documents through typewriting is a tedious task, as the documents are difficult to read and understand. Hence there is a need for a computer-based system to overcome the gap between machines and humans. This project proposes an efficient method for Kannada handwritten character recognition system which uses image preprocessing techniques to enhance the quality of an image and exploring deep learning technique for feature extraction. . The proposed work is concentrated on developing an offline handwriting recognition system to recognize the Kannada statements to a suitable degree of recognition.

Keywords: Kannada character recognition, deep neural networks, OpenCV.

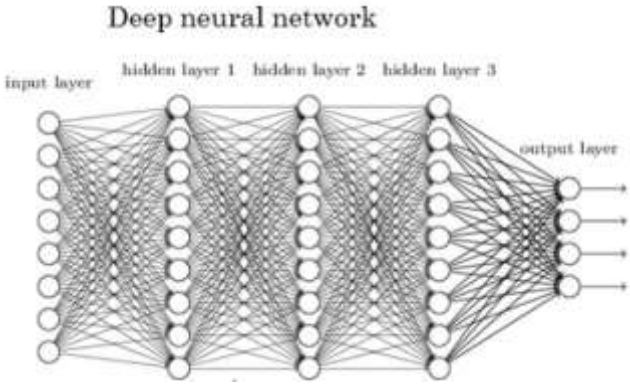
I. INTRODUCTION

Kannada is a Dravidian language predominantly spoken by residents of Karnataka ,which is known to be the official and administrative language. There are numerous old documents and literature which are handwritten in Kannada. Hence there was a need for digitizing these documents for the future use so that there will be knowledge transfer and the past information was preserved. To meet the Kannada publishing needs in the mid-1980s, software development in Kannada was initiated. Use of specific formats for Kannada software development made it highly difficult for data portability across applications since the document was written and saved using one application could not be opened in the other. Maintenance of handwritten documents is highly labor-intensive operation and access of this information is very difficult due to handwriting issues. Hence there is a need for use of technology to store and manage this data, to make it easily accessible and user friendly. Use of Handwritten text recognition system will be a boon to solve this practical issue in government sectors and health departments where most of their documents will be Kannada.

II. ALGORITHM

DEEP NEURAL NETWORK(DNN)

Deep neural network* Nodes are little parts of the system, and they are like neurons of the human brain. When a stimulus hits them, a process takes place in these nodes. Some of them are connected and marked, and some are not, but in general, nodes are grouped into layers. The system must process layers of data between the input and output to solve a task. The more layers it has to process to get the result, the deeper the network is considered. There is a concept of Credit Assignment Path (CAP) which means the number of such layers needed for the system to complete the task. The neural network is deep if the CAP index is more than two.

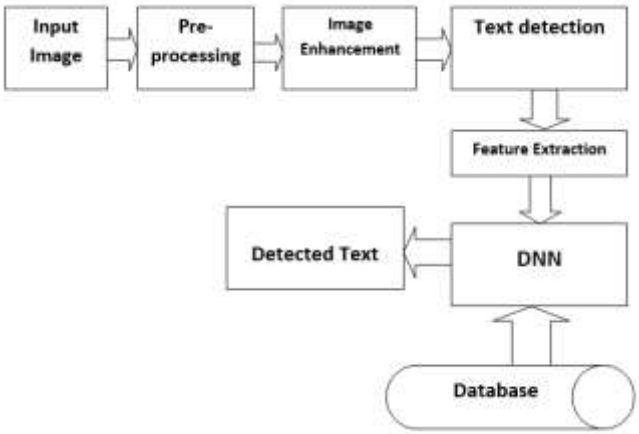


A deep neural network is beneficial when you need to replace human labor with autonomous work without compromising its efficiency. The deep neural network usage can find various applications in real life. For example, The American

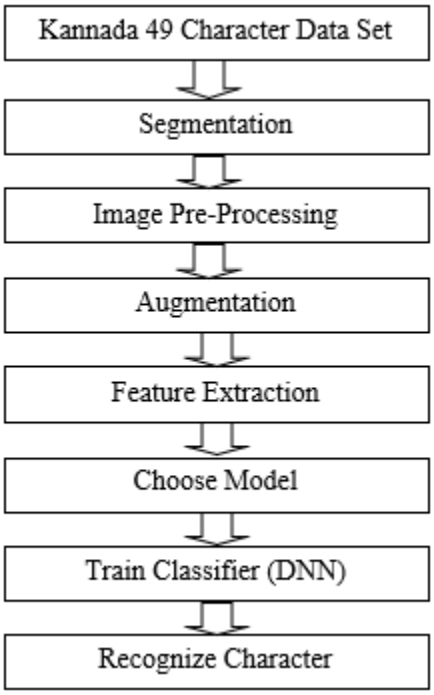
IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 7, July 2021
DOI 10.17148/IJARCCE.2021.10713

company Pony.ai is another example of how you can use DNN. They developed a system for AI cars that can work without a driver. It requires more than just a simple algorithm of actions, but a much deeper learning system, which should be able to recognize people, road signs and other markings like trees, and other important objects.

III. METHODOLOGY



Hand written Kannada text raw image is given as input for this system. Raw image given to preprocessing stage where RGB to grayscale conversion of image will be done. Image quality enhanced by using image enhancement tool box. In this project median filter is used for image enhancing. In this system DNN play main role. Kannada characters are trained and stored in database using Deep Neural Networks. DNN detect text from noisy reduced image and estimate the ROI area of text. Finally Detected texts are extracted using deep neural networks.



Working Principle:

Data Collection: The Kannada 49-character dataset is used for the experimentation purpose. The dataset consists of a collection of images from various sources.

Segmentation: The entire document is first segmented into lines, then words and at last into characters. Contour detection, Dilation and Bounded boxes method has been used for same.

Image Preprocessing: For the development of computer vision algorithms, image pre-processing plays a vital role in enhancing the quality of the image. In the preprocessing step, the input is usually an image and output obtained is the

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 7, July 2021
DOI 10.17148/IJARCCE.2021.10713

characteristics associated with that image. The preprocessing activities include grayscale conversion, noise removal, contrast normalization, binarization, and segmentation.

Augmentation: Since certain models require large number of training data, it is practically not possible to write them. Due to this, we augmented the existing dataset to create more dataset for training phase. This has to be done with care so that over fitting does not happen.

The augmentation steps we considered were: Aspect ratio, rotation, smoothening, padding, noise, resizing.

Feature extraction: For extracting features from handwritten characters, a zoning technique is selected. In this technique, a zoning grid is overlapped on the considered images and the significant feature values are accumulated based on the density of pixels, color or values of pixels in corresponding grids. This technique works on u 3 v grid that divides an image into uniform sub-samples and the resultant sub-samples having identical shapes. These identical shapes know called zones.

Choose Model: Deep Neural Network is chosen in this project.

Train Classifier(DNN): Nodes are little parts of the system, and they are like neurons of the human brain. When a stimulus hits them, a process takes place in these nodes. The system must process layers of data between the input and output to solve a task. The more layers it has to process to get the result, the deeper the network is considered.

Recognition of Kannada Character: Due to the variations that occur in shapes, CNN are used for classification and recognition of objects.

IV. CONCLUSION

Due to Deep Neural Network (DNN), a deep learning technique has effectively extracted the features from the input image which helped the model in better classification of Kannada handwritten character recognition. Categorical Cross-entropy loss function was used to calculate the error rate in the model. For the dataset collected, an accuracy of 95.11% was achieved during training. However, when real-time handwritten data used during testing of the model, an accuracy of 86% was achieved. However, the computational speed could have been improved if the system installed with GPU. The variation in the accuracy is because the handwriting of each individual differs and may lead to the overlapping of letters which might be a reason for reduced accuracy values for real-time handwritten data. Since Kannada handwritten reports were the only way of documentation available in government offices, healthcare departments in Karnataka, reproducing the contents of these old documents through typewriting would be a tedious task which is labor intensive and may lead to errors during documentation. Hence, the proposed model would be highly helpful in bridging the gap between humans and machines and would find its application in government sectors for reproducing the old Kannada documents.

RESULT

1: Initial window process

2**:** Running the program

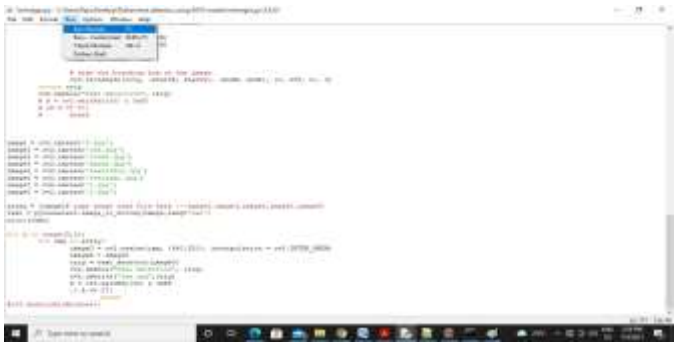
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

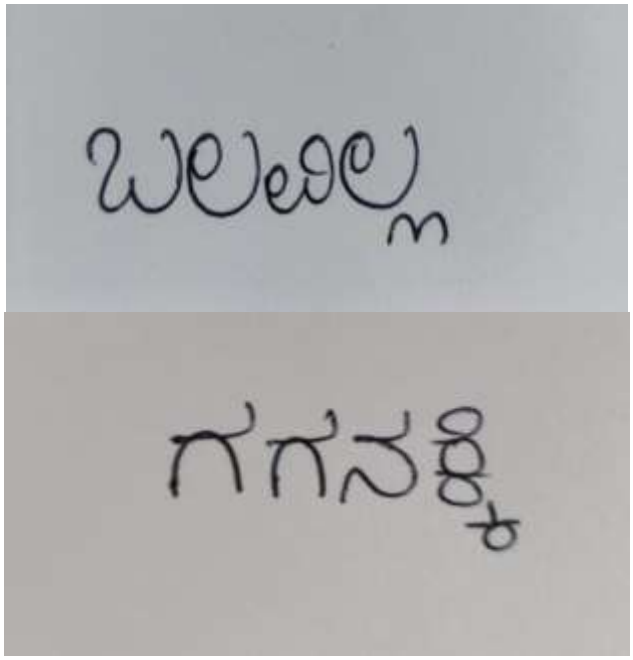
International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

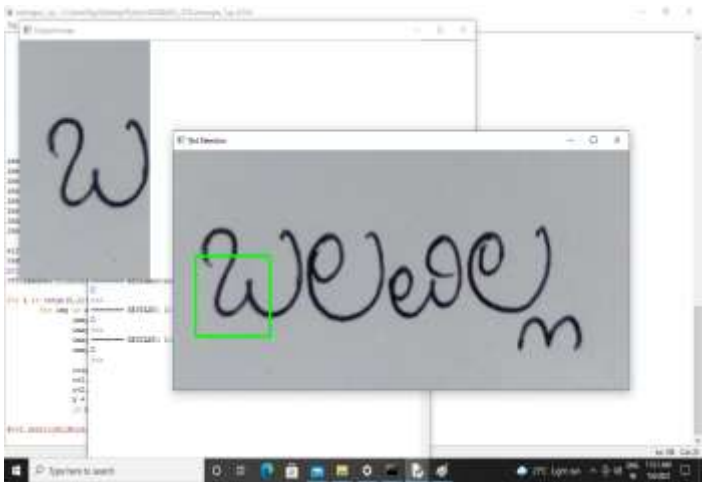
DOI 10.17148/IJARCCE.2021.10713



3: INPUT IMAGES- Feeding the images to search one particular character in the given input word.



4: OUTPUT IMAGES



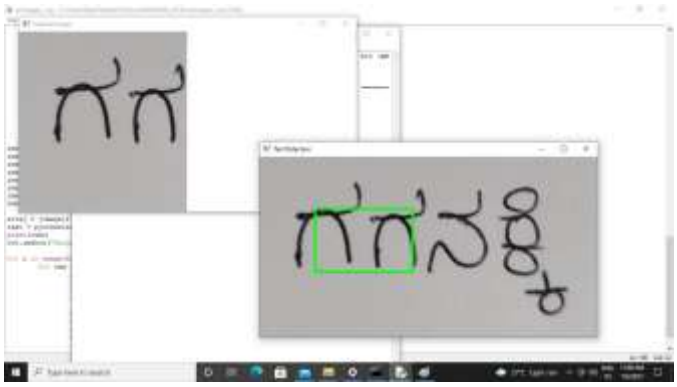
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10713



REFERENCE

[1] S. Khan, H. Ali, Z. Ullah, N. Minallah, S. Maqsood, and A. Hafeez, ''KNN and ANN-based recognition of handwritten pashto letters using zoning features,'' International Journal of Advanced Computer Science and Applications, vol. 9, no. 10, pp. 570–577, 2018.

[2] Sahloul A and Suen C, OFF-line system for the recognition of handwritten arabic character. In: Fourth international conference on computer science & information technology. Dubai, UAE, 2014, pp. 227–244.

[3] J. Hull, "A database for handwritten text recognition research," IEEE Trans. On PAMI, 16(5), 1994, pp. 550–554.

[4] M Liwicki, H Bunke, in Proceedings of the Eighth International Conference on Document Analysis and Recognition. Iam-ondb – an online english sentence database acquired from handwritten text on a whiteboard, (2005), pp. 956–961.

[5] UV Marti, H Bunke, The iam-database: An english sentence database for offline handwriting recognition. Int. J. Doc. Anal. Recognit. **5**(1), 39–46 (2002).

