International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 07 | July -2020 ISSN: 2582-3930

Hand Written Text Recognition using Capsule Networks

[1] PRAVEEN N, [2]TEJASWINI K, [3]JAISHREE S, [4]DEEPIKA R

[1] Assistant Professor, [2] UG Student, [3]UG Student, [3]UG Student

[1] [2] [3] [4]Computer Science and Engineering,, Cambridge Institute of Technology

Abstract - Handwritten letter recognition using computer vision and Machine Learning Technologies has been a well pondered upon topic, since the emergence of Machine Learning, the field has undergone tremendous development. As a subfield of Computer Vision, handwritten text recognition plays an important role in helping the machines to gain humanlike characteristics. There has been many experimentations on various literatures with techniques like Multiscale partial differential operators, Support vector networks, Deep belief networks and Artificial neural networks yielding acceptable accuracies. As convolutional neural network has a strong architecture, inspired by the same architecture and selectively trying to overcome the drawbacks of CNN, we use a new robust and dynamic technology of Capsule neural networks, which yields a considerably high quality results in terms of accuracies and other metrics such as precision and specificity when evaluated.

***Key Words*: Machine learning, Convolutional networks, Capsules, handwritten characters, Kannada**

1.INTRODUCTION

Computer vision is a successful, fast-growing field of artificial intelligence that trains computer to understand and interpret using digital images and deep learning models. Handwritten text recognition has been a widely explored topic in the computer vision field. One of the important challenges of this field is to develop an efficient handwritten text recognition technology. The main aim is to translate old manuscripts, reading sign boards, literature, and assistance to the blind. Digitization of larger amounts of data or documents can be carried out easily with the help of handwritten text recognition technology [1]. Many advancements have been made in this field like identification, tracking, measuring, detecting and classifying objects and images [2].

Deep learning based methods are being used in the resent years. Many other works have been applied in the recent years with similar concepts of CNN [3] to train the neural network. However, with these methods redundancy was the major problem. CNN is dominantly being used in the recognition of other Indian languages like Gurumukhi, Bengali or Davanagari languages [4][5][6]. Some of the advanced techniques such as feature extraction methods are being used in many international languages such as English, and neural networks have been used in the classification of other foreign languages such as Arabic, Chinese, Latin and Japanese.

India being one of the most linguistically diverse country in the world, its diversity is so large that almost every state has its own personal language. Hence, it's very important to recognize these languages. Very minimal efforts have been noted with the introduction to the state of the art technology in the field of handwritten text recognition to Indian regional languages. Majorly of the above work used support vector machine to classify the handwritten characters, some of the recent works based on south Indian language recognition include the use of feature vectors with machine learning for classifying the characters. Further , artificial neural networks was used with various feature extraction techniques with the help of Convolution neural networks (CNN). The CNN consisted of many neural layers which then scaled the features as scalars. Max pooling was backed for a particular feature of interest, but the orientation of the features were lost or misplaced in this process of scaled features called scalars. due to the above problem a new network called capsule networks was introduced to negate the problem of disorientation by considering features as vectors, by having the information about set of instantiation parameter such as pose and probability of that feature, which then combining these two information gives us its true orientation. To ensure that the output of a layer is directed from a particular preceding layer dynamic routing is used [7].

This paper aims on classifying the Kannada handwritten text recognition with the help of classifier techniques. The remainder of this paper is organized as follows. Section I presents the Introduction and section II reviews the related work. Section III describes the problem statement and section IV reveals the architecture. Proposed system is explained in section V. The results are discussed in section VI and we conclude this paper in section VII

2. Related Work

Character recognition is one of application of computer vision which recognises charters from sources like image, hand written patterns, documents and many more. Handwritten character recognition lies in the field of recognition of patterns which defines an ability of a machine to analyze patterns , identify and classifiy the character of different languages. For recognition of character., the text must be converted to editable documents which is done by OCR(Optical Character

Recognition) OCR can extract text from a scanned document or an image of a document. Some of the advantages of OCR are Higher Productivity. Cost Reduction, High Accuracy, Increased Storage Space, Superior Data Security and many more.

India has wide variety of languages in which Kannada langauge is one among them and challenging work is to recognise kannada charcter and also kannada numerals because of the curves in the language., S. Karthik and K. Srikanta Murthy [8] Contributes towards recognition of numericals in kannada using gradient descriptors for feature extraction and SVM for classification. The proposed algorithm is experimented on 4,000 images of singled out handwritten numerals of the kannada language and an whopping accuracy of 95% is achieved.

OCR plays an important role in reading text from image or document is presented by Matteo Brisinello & Ratko Grbić [9] a preprocessing method is presented for improving Tesseract OCR. There are 2 steps firstly, a text segmentation method And secondly, a classifier is used to identify the image containing text among k images.

Here is an another approch for kannada numeral recognition.,

- Vishweshwarayya C. Hallur and R. S. Hegadi [10] introduce a system which is a holistic based approach that means the entirety of the input is tested on a database consisting of around 1470 images as 10 numerals are each written by 147 contributors.

Social media now a days contain number of images that are widely used and shared all around the corners of the world. They need to be recognised and analyse text. MS Akopyan OV Belyaeva and TP Plekov [11] show that the extraction of text is on various images of different qualities is done, then pre-processing is done for input image and later OCR is applied to recognise the text. Gradient boosting neural networks are used for classification in this work.

Neural Network is very extensive and they solve the sample recognition problems and recognition of characters is one among them., Desai Hinduja R. Dheebhika T. Prem Jacob [12] It recognizes character, numbers, special characters. The solution is obtained using MATLAB NN toolbox.

Some of the challenges involved is the font characteristics of the characters in paper documents and quality of images is answered by Pratik Madhukar Manwatkar & Shashank H. Yadav [13] , The objective of this work is to help in the understanding of an user to read without any disruptions by using specific sequences of discrete processing modules.

Hand written character recognition also must be an effective and an in demand technique for recognition of cursive handwritten texts. This work from Shahbaz Hassan, Ayesha Irfan, Ali Mirza & Imran Siddiqi [14] with a case study on Urdu Handwriting goes on with convolutional neural networks (CNNs) that are employed as feature extractors while classification is carried out using a bidirectional Long-Short-Term Memory (LSTM) network. The proposed technique is validated on a dataset of 6000 unique handwritten lines of text.

3. Proposed System

As image classification and detection has become one amongst the key pilot use cases in the field of machine learning and convolutional neural networks are the most popular technique in this field, but as there is no spatial information used, the techniques showcase some snags. As Capsule networks overcome this problem through understanding an image by mimicking the way of the human vision and how the information is processed and it is learnt.

We use this effort in order to recognize and classify a Dravidian Language [15], Kannada, from different sources of written works.

4. SYSTEM ARCHITECTURE

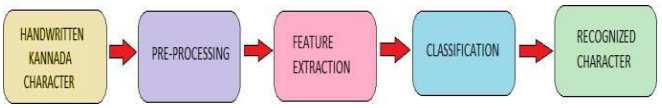


FIG 1 System archtecture

The above figure 1 represent the system architecture ., In this initially a sample kannada **character image** either vowels or consonants is given as input. Next step is **Pre-Processing** the input image ie., converting image to grey scale image in dimension 28x28. The next step or process is **Extracting Features** is a special form of dimensionality reduction and obtaining important features such as curve in the letter. The next process is **classification** using capsule Network and result obtained is recognised kannada hand written character.

5. PROPOSED METHOD

5.1 Brief

Convolutional neural networks are widely used for image processing and detection but there are drawbacks like they only train well with a large number of images as input, they are poor in handling ambiguity and they usually lose a lot of data in the pooling layer henceforth resulting in an output that is usually invariant to minute changes. On the contrary to CNNs capsule technology consists of capsules which imitate the human neurons, a capsule is a little class of neurons that works on detecting a specific object within a region of an image and the output is a vector whose estimated length represents the probability of the presence of the object hence resulting in better results and becoming the state of the art technology.

5.1 Work flow of the capsules

A capsule network is arranged in a fashion of multiple layers. The primary layer, the layer which is at the bottom is where the receiving of a small region of the image happens as shown in the fig [untitled] the part of the letter Ka highlighted in red is fed into the primary layer. The existence and the pattern's pose is discerned.

Fig 2 Region of the image input

The dataset that is used here consists of numerous images of the same alphabet as shown in the fig to ensure a better training of the network as the patterns may differ in distinct writings.

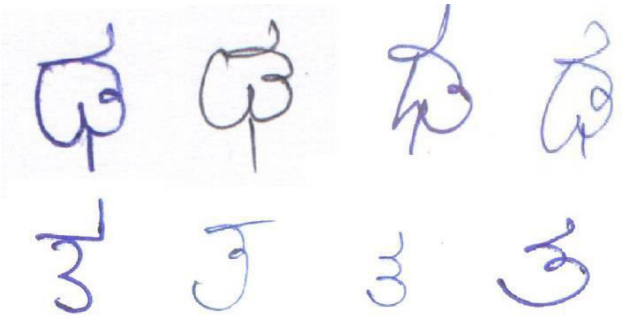


Fig 3 Illustrative of the Data set.

As capsule networks differ from CNNs by ditching the max pooling method and avoids losing valuable information by encapsulating all information about the state of the feature they are detecting in vector form, encoding the relative spatial relationships between features and the probability of detection of a feature as length of their output vector.

The Primary layer is made up of some convolutional layers, ass own in the figure 4 the input layer which consists of pre-processing which is followed by two convolutional layers which outputs feature maps consisting of scalars. These scalars are then converted into dimensional vectors in the capsule's Reshape layer. Squash layer whose functions guarantee to make the length of the vector in between 0 and 1 result in the final probabilistic value.

The following layers after the primary layer is where the routing algorithm known as the routing by agreement works. The product of capsules is the dimensionality of the output vector and with n number of classes fed as input gives out the character capsule layer's structure.

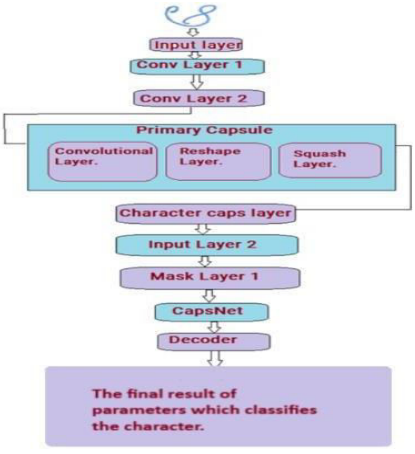


Fig 4 Flow of the detection from capsule networks

The same output is taken from the layer and is fed into the decoder network. The mask layer 1 which is a part of decoder network and maximises the capsule output with highest vector length and feeds this input to the decoder model. The dense layers of the decoder network are fully connected neurons and every output from the layer preceding is weighted and directed to the next neuron. The layers of the decoder network collectively decode the input vector and perform the classification.

6. RESULTS AND DISCUSSION

In this research work experiments were carried out with database containing 23500 dataset. The hand written texts have been collected from different age groups. As we discussed kannada alphebets is case sensitive because of the curves in langauge. The alphabet is trained that is as shown below in Test cases. TensorFlow is a large library, and depending on the full package when writing a unit test for its submodules has been a common practice. Basically ,software testing, we look out for bugs and mistakes that may result in unexpected features or system crashes and there are chances of even for machine learning models . it involves the clean structural design and careful implementation of models, layers, or basic mathematical operations.Therefore we have conducted software testing using tensorflow. Unit testing is suitable in this when we want to test the correctness of our own implementation. Tensorflow unit test is very similar to the python unit test. The test cases are performed both on Vowels and consonants as shown below.,

6.1 Test cases for Vowels

Case1:

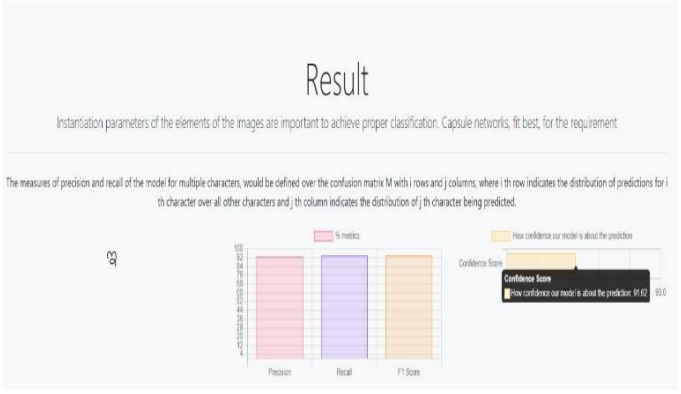


FIG 5 Represent letter [E]

Letter: [E]

Confidence score : 93.0%

Case 2:

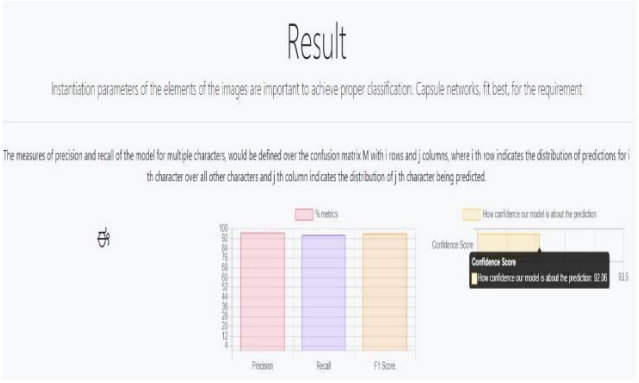


FIG 6 Represent letter [EE]

Letter: [EE] 

confidence score : 92.05%

Case 3

FIG 7 Represent letter [AaH]

Letter : [Aah] 

Confidence score : 93.15%

Case 4

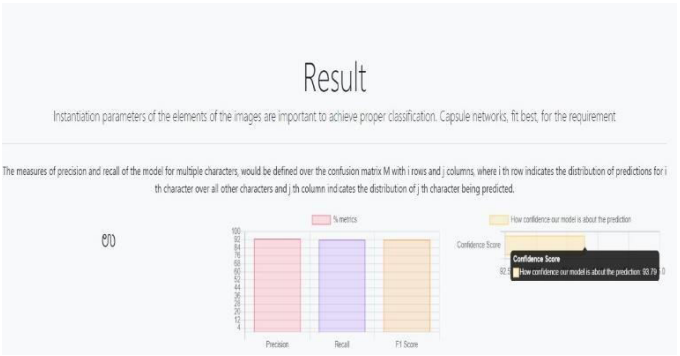


FIG 8 Represent letter [OOH]

Letter : [ooh] 

Confidence score : 93.79%

Case 5

FIG 9 Represent letter [OOOH]

Letter : [oooh] 

Confidence score : 94.84%

6.2 Test cases for Consonants

Case 1

FIG 10 Represent letter [Kha]

Letter :[Kha] 

Confidence score : 96.65%

Case 2

FIG 11 Represent letter [Ga]

Letter :[Ga] 

Confidence score : 91.77%

Case 3

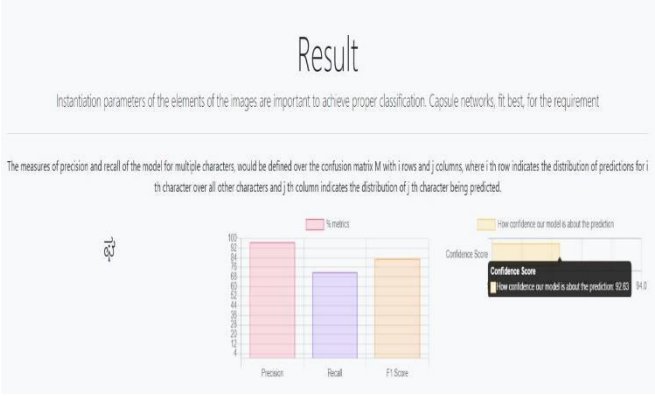
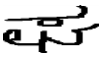


FIG 12 Represent letter [Gha]

Letter [gha] 

Confidence score : 92.63%

Case 4

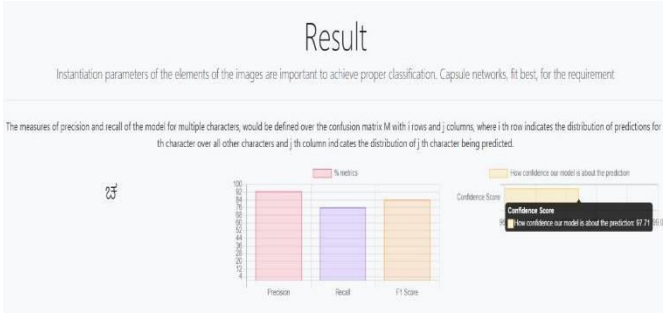
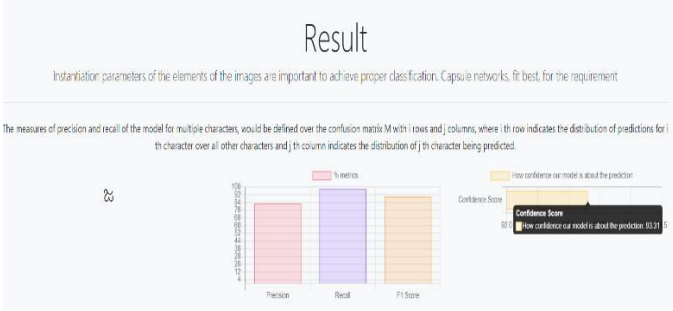


FIG 13 Represent letter [Cha]

Letter : [cha] 

Confidence score : 97.71%

Case 5



Letter : [jha] 

Confidence score : 93.31%

The above represents the different accuracies for different kannada alphabets. These accuracy gives an insight of model`s performance. The accuracy can be calculated by formulae:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

The difference in accuracies are explained by the structural similarities between the consonants and the accuracy estimated for vowels and consonants are 98.70 and 98.70 respectively for training and 94.95 and 83.94 respectively for validation score.

7. Future work

As we discussed from related work [9] focuses on pre-processing of data , [14] uses different classification technique for feature extraction and classification ., this could enhance the performace of our model also.

- For an existing model , maxpooling layers can be an addition to improve the current existing system as of theoritically.
- Histogram equalization can be applied for image enhancement ., this could expect improvment in the accuracy of the model , but this is applicable for grey scale image and given as input for first convolutional layer as of theoritically

REFERENCES

1. M. Acharya, P. Chouhan and A. Deshmukh, "Scan.it - Text Recognition, Translation and Conversion," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 2019, pp. 1-5.
2. Y. Zhu, C. Yao, and X. Bai, “Scene text detection and recognition: Recent advances and future trends,” Frontiers of Computer Science, vol. 10, no. 1, pp. 19–36, 2016.
3. N. Jmour, S. Zayen and A. Abdelkrim, "Convolutional neural networks for image classification," 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, 2018, pp. 397-402.
4. G. S. Lehal, C. Singh and R. Lehal, "A shape based post processor for Gurmukhi OCR," Proceedings of Sixth International Conference on Document Analysis and Recognition, Seattle, WA, USA, 2001, pp. 1105-1109.
5. Majumdar A., Chaudhuri B.B. (2006) A MLP Classifier for Both Printed and Handwritten Bangla Numeral Recognition. In: Kalra P.K., Peleg S. (eds) Computer Vision, Graphics and Image Processing. Lecture Notes in Computer Science, vol 4338. Springer, Berlin, Heidelberg.

6. M. Jangid, R. Dhir, R. Rani and K. Singh, "SVM classifier for recognition of handwritten Devanagari numeral," 2011 International Conference on Image Information Processing, Shimla, 2011, pp. 1-5.

7. S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, pp. 3856–3866.

8. S. Karthik and K. S. Murthy, “Handwritten kannada numerals recognition using histogram of oriented gradient descriptors and support vector machines,” pp. 51–57, 2015.

9. M. Brisinello, R. Grbić, D. Stefanovič and R. Pečkai-Kovač, "Optical Character Recognition on images with colorful background," 2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), Berlin, 2018, pp. 1-6.

10. V. C. Hallur and R. Hegadi, “Offline kannada handwritten numeral recognition: Holistic approach,” pp. 632–637, 2014.

11. M. S. Akopyan, O. V. Belyaeva, T. P. Plechov and D. Y. Turdakov, "Text Recognition on Images from Social Media," 2019 Ivannikov Memorial Workshop (IVMEM), Velikiy Novgorod, Russia, 2019, pp. 3-6.

12. D. Hinduja, R. Dheebhika and T. P. Jacob, "Enhanced Character Recognition using Deep Neural Network-A Survey," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0438-0440.

13. P. M. Manwatkar and S. H. Yadav, "Text recognition from images," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2015, pp. 1-6.

14. S. Hassan, A. Irfan, A. Mirza and I. Siddiqi, "Cursive Handwritten Text Recognition using Bi-Directional LSTMs: A Case Study on Urdu Handwriting," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019, pp. 67-72.

15. Kolipakam, V., Jordan, F. M., Dunn, M., Greenhill, S. J., Bouckaert, R., Gray, R. D., & Verkerk, “A bayesian phylogenetic study of the Dravidian language family”. *Royal Society Open Science*, *5*(3), [171504]. 2018.