

Siamese Neural Network Model for Recognizing Optically Processed Devanagari Hindi Script

Ambili P S . Agnesh L . Arun K V

REVA University, Bengaluru, India

Received: 02 May 2023 / Revised: 18 May 2023 / Accepted: 06 June 2023
©Milestone Research Publications, Part of CLOCKSS archiving
DOI: 10.5281/zenodo.8210372

Abstract – Optical Character Recognition (OCR) is a technology that allows computers to recognize and extract text from images or scanned documents. It is commonly used to convert printed or handwritten text into machine-readable format. This Study presents an OCR system on Devanagari script Hindi Characters based on Siamese Neural Network. Here the Siamese Neural Network, a Deep neural network which comprises of two identical Convolution Network compare the script and ranks based on the dissimilarity. When lesser dissimilarity score is identified, prediction is done as character match. In this work the authors use 5 classes of Hindi characters which were initially preprocessed using grey scaling and convert it to pgm format. This is directly input into the Deep convolutional network which are learnt from matching and non-matching image between the CNN with contrastive loss function in Siamese architecture. The Proposed OCR system uses very less time and gives more accurate results as compared to the regular CNN.

Index Terms – Optical Character Recognition, Deep learning, Siamese Neural Networks, Deep neural Network, Dissimilarity Score.

I. INTRODUCTION

Optical Character Recognition (OCR) is a technology that enables the conversion of handwritten or printed text into digital format that can be processed and edited by computers[1]. It has been widely used in various applications, such as document digitization, language translation, and information retrieval. However, developing an OCR system for non-Latin scripts, such as Hindi, is challenging due to the complexity and diversity of the script. Siamese Neural Networks (SNNs) have been used for various applications, including face recognition, signature verification, text similarity, image retrieval, document matching, speech recognition, recommendation systems, and bioinformatics[2,4,5]. This Research shows the development of Hindi OCR system that can recognize the Hindi Swar with high accuracy and precision characters which was trained. With the advancement of deep learning techniques and the availability of large datasets, SNNs are becoming increasingly popular in various industries, including finance, healthcare, and entertainment [15].

Research Gap Analysis: Many Research and experiment of OCR is carried forward with Convolutional Network and RNN which has an upper hand towards finding patterns in images but it takes lot of data and computation resource for training the model, Siamese model uses far less data than that of traditional CNN and RNN and very few research has been done with OCR with Siamese Network, also since Hindi as a language has curves in its writing style Siamese network excels in those front too[17,18].

The key steps in the proposed methodology are:

1. Collect and pre-process the data, including image cropping, resizing, and normalization. In the case of OCR, this may involve converting text images to grayscale or binary format.
2. Create a dataset of image pairs, where each pair contains two images of the same or different characters. Label the pairs as positive(1) or negative(0), depending on if it is the same character then that pair will be labelled 1 and if they are having different characters then that pair will be labelled 0.
3. Define the Siamese neural network architecture with two identical networks that share the same set of weights. Both the network architecture works are based on Convolution Neural Networking [2,6,10]. Two images are passed as input at the same time to both branches of the Siamese Networks and then are made to train on it and at the end the output of both the networks are compared using distance or similarity metrics.
4. Use of contrastive loss function to reduce the distance between positive data pairs and increase the distance between negative data pairs.

II. LITERATURE REVIEW

Siamese neural networks have been extensively studied for natural language processing tasks, such as sentence similarity, paraphrase identification, and text classification. Optical Character Recognition (OCR) is an important application of these networks, but there is limited research on applying them for Hindi character recognition[7]. Convolutional neural network has been the go to model for OCR for any lingual as they learn in a hierarchal form which is helpful in recognizing the image or the characters. Other methods such as RNN and LSTM also have been experimented with OCR where LSTM used to handle long term dependency which also fairs well with cursive characteristics of a character [12,13]. Convolutional Neural Networks (CNNs) are highly effective for image recognition tasks, but they have some disadvantages. These include complexity, over fitting, lack of interpretability, difficulty with variable-sized inputs, limited context awareness, limited rotation and translation invariance, and limited object orientation and position. CNNs are designed to be invariant to small translations in the input image but may not be as effective at handling large translations or rotations [8].

Hindi as a language while making a OCR have a potential disadvantages with CNN, RNN as they Lack of training data, difficulty with complex scripts, dependence on pre-processing, difficulty with handwriting recognition, and computationally expensive can make it difficult to deploy an OCR system in real-time or on resource-constrained devices [3,9,14]. Kannada is a complex script with multiple

components or diacritics, making it difficult to design a CNN that can effectively recognize all these variations. Handwriting is highly variable and can be difficult to model effectively with a CNN.

Despite the variety of models and techniques used in previous research, there is more sophisticated approaches which was explained in this study where Siamese Neural Networks which is used where the network consists of two identical sub-networks, often called "twin" networks that share the same weights and architecture [1]. Each sub-network takes in one input image and processes it through a series of convolutional and pooling layers to extract relevant features [3]. The extracted feature maps are then flattened and fed through a series of fully connected layers that output a feature vector for each input. Overall, this methodology require fewer training samples than CNNs due to their similarity or distance learning. They can be adapted to new tasks by fine-tuning a pre-trained network on the new data [11,14].

III. METHODOLOGY

Siamese networks can be used to compare two different images of the same character or word and determine whether they represent the same or different characters/words [19]. This can be useful in scenarios where the input data is noisy or has variations in style or orientation. One of the main advantages of Siamese networks is that they can be trained with relatively small amounts of labelled data, making them useful in applications where large datasets are not available. Additionally, Siamese networks can be trained end-to-end, allowing them to learn features directly from the input data without the need for hand-engineered features.

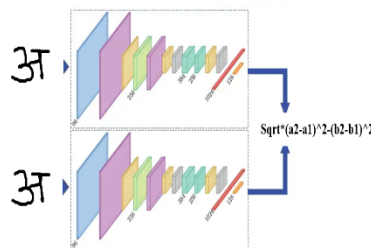


Figure. 1: Proposed Architecture of Siamese Network

Once the feature vectors have been generated, a distance metric is computed between them to determine their similarity. This distance metric can be computed using various techniques such as Euclidean distance, cosine distance, or contrastive loss function. This network takes in two inputs and computes a similarity score between them, which can be used to determine if they belong to the same class or not. The Siamese network is trained using a contrastive loss function, which helps to optimize the network's parameters for better accuracy. Adam optimizer is a gradient-based optimization algorithm used in deep learning models to update the weights of the neural network during training [16]. Adam optimizer uses the first and second moments of the gradients to estimate the adaptive learning rates for each parameter in the neural network. It works by computing the current gradient of the loss function with respect to the parameters, then updating the parameters in the opposite direction of the gradient, with a learning rate

proportional to the first moment divided by the square root of the second moment. It can handle sparse gradients and noisy data, has low memory requirements, and is computationally efficient. It also has automatic learning rate tuning, which helps to converge faster and more accurately to the global minimum.

The contrastive loss function (clf) is a loss function used in the training of Siamese neural networks to measure the similarity between two inputs. It takes two inputs and a label as input and penalizes the model if the distance between positive pairs is greater than a margin value and the distance between negative pairs is smaller than the margin value.

$$clf = yd^2 + (1-y) \max(\text{margin} - d, 0)^2 \tag{1}$$

Where ‘y’ is the label and ‘d’ is the distance between the two inputs.

‘d’ is calculated as

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2}$$

The label y is 1 if positive pair and 0 if negative pair. If the inputs are a positive pair (y=1) and if the labels are of the same class (positive pair) or different classes (negative pair), the loss function penalizes them if the margin value is greater than 0.

IV. RESULTS AND DISCUSSION

Siamese neural network unlike other classification and regression models which gives out classes and continuous values as output but this network provides the distance between the given input by considering the distance between the input class and the class which is been passed. The output is based on the similarity and dissimilarity values which is achieved the lesser the dissimilarity score it is likely that the input value is same as the class. The study used Hindi character set of 85690 characters belonging to approximately 92 classes with 305-590 sample images in each class. The study classified five classes of Hindi characters using Siamese Neural Network with almost 420 sample images in each class. The similar swaras showing low dissimilarity index and dissimilar swaras identified correctly based on high dissimilarity index are presented in Figure.2 and Figure.3 respectively.

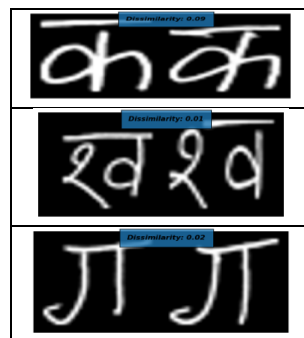


Figure.2: Similar Swara showing low dissimilarity.

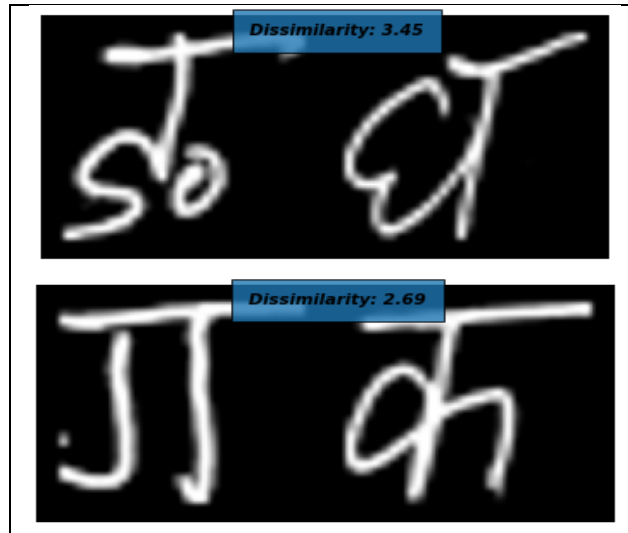


Figure.3: Dissimilar swaras showing high dissimilarity.

Evaluation and Deployment of the model

The trained model is evaluated by separate test set which checks the accuracy and performance by the metrics of precision, F1-Score and Recall, Dissimilarity [6,9]. The deployment of the model was done on Kivy tool used for user interface creation and it works well on python language [8].

Major advantages of proposed system are:

1. Effective identification of handwritten and Optically processed Hindi scripts.
2. One of the most precise models in deep learning by using two CNNs to identify accurate dissimilarity between the input characters.
3. The model can become a powerful tool for identification, particularly in situations where there is a high degree of variation in writing styles or limited training data is available.

The training and validation accuracy are presented in Figure 4 and training and validation loss graphs is Figure. 5 .



Figure.4: Accuracy and validation accuracy

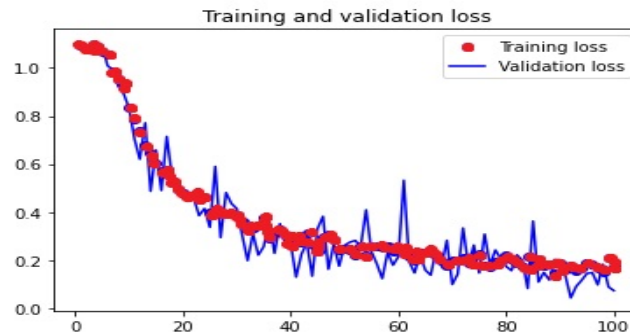


Figure. 5: Training Loss and validation loss

V. CONCLUSION

This Research illustrates a Siamese neural network (SNN) with custom training methodology on a pytorch framework. It successfully classifies all the trained classes of Hindi characters and gives the dissimilarity of each class with respect to the newly passed image. The developed SNNs have been shown to be effective in handwritten character recognition, where there is a high degree of variation in writing styles. This methodology also supports when data to be trained are very diverse which are true for handwritten texts. Also fairs well compared to the RNN and CNN which uses more computation, training data and memory as compared to the Siamese network. In short Siamese network can be fitting to made a OCR model and also gives promising results.

REFERENCES

1. Kaur, J., Goyal, V., & Kumar, M. (2021, November). Tesseract OCR for Hindi Typewritten Documents. In 2021 Sixth International Conference on Image Information Processing (ICIIP) (Vol. 6, pp. 450-454). IEEE.
2. Melekhov, I., Kannala, J., & Rahtu, E. (2016, December). Siamese network features for image matching. In *2016 23rd international conference on pattern recognition (ICPR)* (pp. 378-383). IEEE.
3. Barakat, B. K., Alasam, R., & El-Sana, J. (2018, April). Word spotting using convolutional siamese network. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)* (pp. 229-234). IEEE.
4. Naseer, A., & Zafar, K. (2022). Meta-feature based few-shot Siamese learning for Urdu optical character recognition. *Computational Intelligence*, 38(5), 1707-1727.
5. Sokar, G., Hemayed, E. E., & Rehan, M. (2018, November). A generic OCR using deep siamese convolution neural networks. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 1238-1244). IEEE.
6. Ahmed, S. T., & Basha, S. M. (2022). *Information and Communication Theory-Source Coding Techniques-Part II*. MileStone Research Publications.
7. Jenckel, M., Bukhari, S. S., & Dengel, A. (2018, August). Transcription free lstm ocr model evaluation. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 122-126). IEEE.
8. Jain, M., Mathew, M., & Jawahar, C. V. (2017, November). Unconstrained OCR for Urdu using deep CNN-RNN hybrid networks. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)* (pp. 747-752). IEEE.
9. Mathew, M., Singh, A. K., & Jawahar, C. V. (2016, April). Multilingual OCR for Indic scripts. In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)* (pp. 186-191). IEEE.

10. Ambili, P. S., & Abraham, B. (2022). A Predictive Model for Student Employability Using Deep Learning Techniques. *ECS Transactions*, 107(1), 10149.
11. Avadesh, M., & Goyal, N. (2018, April). Optical character recognition for Sanskrit using convolution neural networks. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)* (pp. 447-452). IEEE.
12. Sharma, A. S., Mridul, M. A., Jannat, M. E., & Islam, M. S. (2018, September). A deep cnn model for student learning pedagogy detection data collection using ocr. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-6). IEEE.
13. Jain, M., Mathew, M., & Jawahar, C. V. (2017, November). Unconstrained OCR for Urdu using deep CNN-RNN hybrid networks. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)* (pp. 747-752). IEEE.
14. Jatowt, A., Coustaty, M., Nguyen, N. V., & Doucet, A. (2019, June). Deep statistical analysis of OCR errors for effective post-OCR processing. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 29-38). IEEE.
15. Alla, K., & Prasad, R. S. R. (2010, April). A new approach to Hindi text steganography using matraye, core classification and HHK scheme. In *2010 Seventh International Conference on Information Technology: New Generations* (pp. 1223-1224). IEEE.
16. Abraham, B., & Ambili, P. S. (2023). An Enhanced Career Prospect Prediction System for Non-computer Stream Students in Software Companies. In *Computational Intelligence for Engineering and Management Applications: Select Proceedings of CIEMA 2022* (pp. 811-819). Singapore: Springer Nature Singapore.
17. Bansal, V., & Sinha, M. K. (2001, September). A complete OCR for printed Hindi text in Devanagari script. In *Proceedings of Sixth International Conference on Document Analysis and Recognition* (pp. 0800-0800). IEEE Computer Society.
18. Shah, R., Gupta, M. K., & Kumar, A. (2021, November). Ancient Sanskrit Line-level OCR using OpenNMT Architecture. In *2021 Sixth International Conference on Image Information Processing (ICIIP)* (Vol. 6, pp. 347-352). IEEE.
19. Xiao, W., & Wu, D. (2021, December). An improved siamese network model for handwritten signature verification. In *2021 IEEE International Conference on Networking, Sensing and Control (ICNSC)* (Vol. 1, pp. 1-6). IEEE.