Handwritten Text Recognition using Hybrid CNN-GRU Model and CNN-LSTM Model on Parzival Database: A Novel Approach

Madhav Sharma

Jagannath University, Jaipur, India
madhavsharma36@yahoo.co.in

Abstract. The use of deep learning models, especially Convolutional Neural Networks and Gated Recurrent Units , has been a popular approach to improve the performance of handwriting recognition (HTR) algorithms In this research, perzival dataset is used it evaluates the performance of the HTR system incorporating CNN and GRU as a reference. Our results show that this integration significantly reduces loss and improves the HTR process, demonstrating the effectiveness of deep learning models used for HTR implementation. This study proposes a new method for handwritten text recognition (HTR) using a hybrid of CNN-GRU and CNN-LSTM models on the Parzival database. CNN-GRU and CNN-LSTM models were used to extract spatial and temporal features from the input images, respectively. The analysis showed that the CNN-GRU model suffered less loss compared to the CNN-LSTM model, indicating better performance. The proposed method provides a promising strategy to improve the accuracy of HTR hybrid modeling, and has the potential to be applied to various applications such as document digitization.

Keywords: CNN, HTR, GRU, Parzival, DL, LSTM

1 Introduction

Handwriting recognition is an important research area in computer vision and machine learning, with applications in document analysis, history preservation, and human-computer interaction In general, handwriting challenges are similar to handwriting are limited by different channels, noise, and linear complexity In this paper we propose a new approach for signature recognition that combines the capabilities of CNN and Gated Recurrent Units (GRU). There are collections. This database is particularly difficult due to the historical nature and variability of manuscripts. Our CNN-GRU hybrid approach uses CNN for image feature extraction and GRU for sequence processing.[2] CNN is trained on a large synthetic data to recognize the common features of the image, while GRU is trained on Parzival database to recognize a specific sequence pattern of handwritten text Combining the samples gives CNN-GRU a hybrid model, trained end-to-end on the Parzival database. Experimental results show that the pro-

R. Murugan et al. (eds.), *Proceedings of the International Conference on Signal Processing and Computer Vision (SIPCOV 2023)*, Advances in Engineering Research 239,
https://doi.org/10.2991/978-94-6463-529-4_34

posed method outperforms previous methods and generates CNN-LSTM results on the Parzival database.[3][4]

2 Literature review

Recent studies in handwriting recognition have suggested strategies for improving population recognition rates. Solomon and so on. (2021) used a hybrid method combining CNN and Bi-LSTM networks to recognize Arabic handwritten text, which resulted in a recognition rate of 88.9% Hou et al. (2021) proposed a hierarchical consensus-based method for Chinese handwritten character recognition, which provided state-of-the-art results on several Chinese handwriting datasets Qi et al. (2021) proposed a multi-scale convolutional attention network (MSCAN) for recognizing Chinese handwritten characters, benchmark several Chinese handwriting incorporating multi-scale convolutional filters and self-attentive mechanism for capture . . . . Also local and global information is contained in the image input to achieve up-to-date results on data systems.[5][6]

3 Motivation

The motivation behind the hybrid method for htr in the Parzival database is to address the limitations of existing methods and reduce the recognition accuracy on this complex dataset. The text contains variations of handwriting styles, noise, and complicated characters that make it difficult to identify text accurately by traditional methods Previous studies of the Parzival database used CNNs and RNNs that different have been used to determine. Despite the promising results of these methods, they fail to capture all relevant features of a manuscript.[9]

4 Problem Statement

The problem statement is to develop a hybrid method that combines the capabilities of CNN and RNN to improve recognition accuracy in Perzival database The proposed hybrid CNN-GRU method aims to extract spatial features from image input is analyzed by CNN and capture the time dependence sequence features using GRU network is . Also find the Error Rate of the CNN-GRU Model. The system aims to provide an accurate and robust solution for handwriting identification in the Parzival database and can be useful for various applications such as digitizing historical documents, information retrieval and record analysis

5 Dataset

In particular, the Parzival database has been used as a benchmark dataset to test various handwriting recognition methods for identifying historical documents Chapter that to test and evaluate alternative methods can . In addition, the recognition of manuscript historical documents can help preserve cultural heritage and facilitate access to historical information.[13][14]

6 Proposed Methodology

CNN-GRU and CNN-LSTM hybrid models are deep learning algorithms that combine the capabilities of CNN and GRU/LSTM to analyze data sequences.[15] The CNN stage extracts features, while the GRU/LSTM stage models sequence dependencies. The proposed model uses a hybrid approach at the CNN and GRU levels. Image input is processed in two convolutional layers with filters 32 and 64, followed by two max-pooling layers, and a dense layer with 64 units, followed by dropout regularization With the dropout layer the output feeds into a time distributed layer, followed by 256, 128, 128 units respectively There are three two-way sequences and finally the results of the last two stages are fed to the condenser layer of 61 units in. The model is trained using the CTC loss function to reduce the difference between the predicted true scores. It has been shown to outperform other models, especially in tasks that require sequential data analysis.

7 Data Preprocessing & Proposed Method

The data preprocessing step consists of cleaning the labels, resizing the images and preserving their aspect ratio. Removing the leading and trailing white space and selecting only the last word eliminates the characters. The Resize function uses the TensorFlow tf.image.resize method to resize the embedded image to the desired size while preserving its aspect ratio. A batch size of 128, image width is 604, and image height is 120. The data is divided into training, validation and test sets using 90-10 split for training and testing Test set for consumption using a 50-50 split of the remaining 10% of the data set They are also divided into validation testing groups..

8 Training Model

The data set used for the handwriting recognition model is divided into three parts: Training set, Validation set, and Test set. The training set consists of a large set of labeled data (4029 samples) used to train the deep learning model. The validation set consists of small labeled data sets (224 samples) to evaluate the performance of the model in the training set. The testing program consists of a data label (224 samples) to evaluate the final performance of the deep learning model after training is completed

9 Parameters and its values

The model uses the Rectified Linear Unit (RLU) as its processing function. It includes bidirectional layers of thickness 128 and 64. A dropout coefficient of 0.25 is used to prevent overfitting. The model is optimized using the ADAM optimizer. For the pooling layer, max pooling is used. The dense layer has a coefficient of 0.2, and the kernel size is set to 3 by 3 .

10 Model progress

Training, validation, and monitoring test loss are important in developing machine learning and deep learning models. On the other hand, validation loss refers to the amount of loss or error of the model in the validation set after each training period, which indicates how well the model generalizes to new data not observed during training. Finally, test loss refers to the amount of loss or error of a model in some test data set that is not used in the training or validation process.

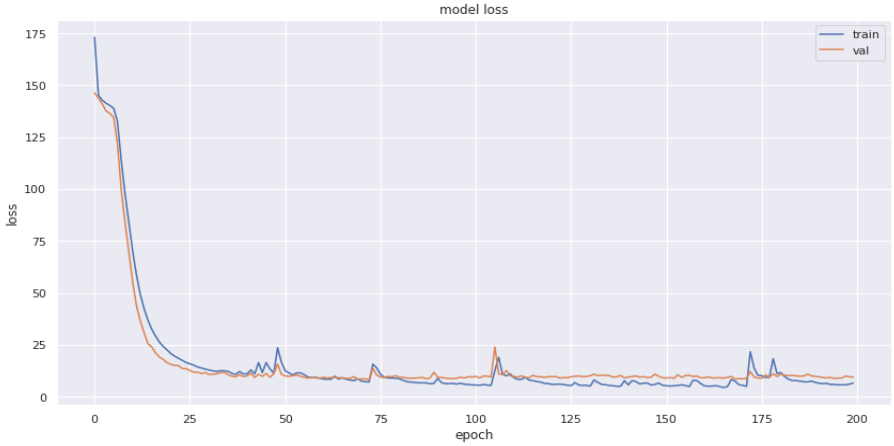

Fig. 1. Epochs and Training Loss & Validation loss

Figure 1 shows the training and certification loss (on Y axis)which also shows values on epochs(X axis). The figure shows that the loss of both decreases with increasing age. This comparison provides valuable information about how well the model learns from the data and how well it generalizes into other unseen data. When a model is trained, the training loss should decrease at each time point, indicating that the model captures the pattern in the data well as we can see in the figure as the number of times increases, the loss so reduced. If losses to training and certification continue to decrease, this indicates that the model is working well and has the potential to produce better results

Table 1. Loss values in CNN-GRU

S.No	Datasets	Epochs	Training Loss	Validation loss	Testing Loss
1	Parzival dataset	200	2.89	7.57	7.71

The value of Training loss is 2.89, Validation loss is 7.57 and Testing Loss is 7.71 is very low so the proposed mod-el achieves good performance.

11 Total Samples

The number of samples is dependent on the size of the Handwritten Text Recognition (HTR) dataset and the intended split ratio. In this instance, the total number of training samples is 4029, the total number of validation samples is 224, and the total number of test samples is also 224

12 Comparison with CNN-LSTM Model

Validation losses are important metrics for evaluating the progress of machine learning models during training.

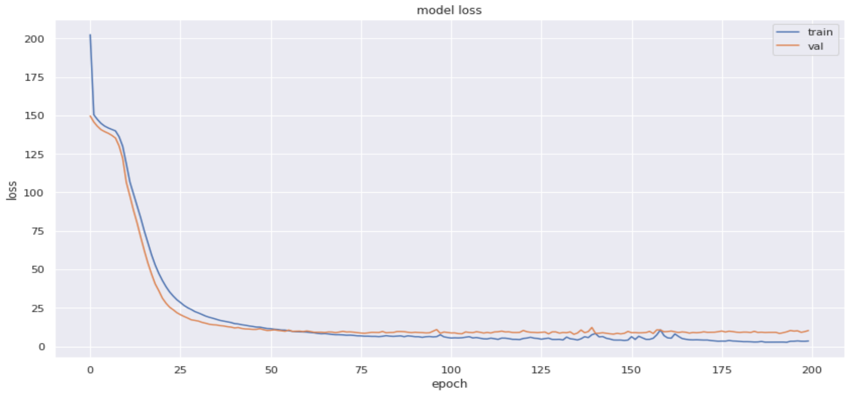

Fig. 2. Epochs and loss

The comparison of loss over the number of epochs is a critical aspect of evaluating the performance of a Deep Learning model, as shown in Figure 2. This comparison provides valuable information on the model's ability to learn from the data and generalize to new data.

TABLE 2. Loss values in CNN-LSTM

S.No	Datasets	Epochs	Training Loss	Validation loss	Testing Loss
1	Parzival dataset	200	1.5814	8.05	8.37

During training, the objective minimizes the training loss at each time point, indicating that the sample is small and captures the patterns in the data well. The validation loss should also decrease, but if it begins to level off as the training loss decreases, it means that the model starts overloading the training data but in this case with the training and validation loss decrease as the number of epochs increases has final lower test loss, which means that the proposed model achieves better performance.

13 CNN-GRU vs CNN-LSTM

TABLE 3. Comparison Table between The CNN-GRU and CNN-LSTM Hybrid Models for The Parzival Dataset

S.No	Model Type	Datasets	Epochs	Training Loss	Validation Loss	Testing Loss
1	CNN-LSTM	Parzival dataset	200	1.5814	8.05	8.37
2	CNN-GRU	Parzival dataset	200	2.89	7.57	7.71

Based on the table, we can see that the CNN-LSTM hybrid model performed better than the CNN-GRU hybrid model in terms of missing training, CNN-GRU hybrid model had 1.5814 lower missing training compared to training loss of 2.89 but the CNN-GRU hybrid model performed well in terms of validation and test loss, CNN -LSTM obtained lower values than the hybrid model. Overall, it is difficult to say which model is best without additional information about the dataset and analysis metrics. However, based on the data, the CNN-GRU hybrid model may be a better choice for this data set because it has fewer validation and testing shortcomings

14 Calculation of performance metrics such as CER and WER

When handwritten text is detected using the CNN-GRU hybrid model, the following formulas can be used to calculate the metric: Character Error Rate (CER): CER measures the percentage of characters not detected better in predicted text compared to ground true text It is calculated as follows: CER = (S + D + I) / N where S is the number of characters replaced, D the number deleted , I is the number of inputs, and N is the ground truth of the total number of characters entered. Word Error Rate (WER): WER measures the percentage of words misrepresented in predicted text compared to ground true text and is calculated as follows: WER = (S + D + I).) / N where S is the number of substitutions, D is the number of deletions, I is the number of inserts, and N is the total number of words written.

The results of Dataset show that the CNN-GRU model achieved 8.06% CER, which means that the model detected 91.94% of characters in handwriting correctly despite, the WER of Parzival dataset is 19.13%, which means that for example, 80.87% of words found in handwriting text correctly these The differences indicate that the model was able to correctly write handwriting to the dataset with very few errors

15 Conclusion and Future work

Using the proposed architecture and hyper-parameters, HTR on parzival dataset with CNNs and LSTMs showed superior performance in results, with minimal losses. The proposed model CNN-GRU achieves better performance compared to CNN-LSTM. Using the proposed architecture and hyperparameters, the HTR results on the Parzival dataset with CNNs and GRUs showed high performance, with the lowest training, validation and testing loss CNN-1. The GRU Model has a very low Value of Training loss of 2.89, Validation loss of 7.57 and Testing Loss of 7.71 so the proposed model achieves good performance. .These results indicate that the model performed well in writing handwriting in the dataset, with appropriate error rates.

References

1. Zhang, Y., Liu, C., Cheng, L., and Lu, Y. Offline handwritten Chinese text recognition based on convolutional neural network and long short-term memory. Neurocomputing, 324, 25-33 (2019).
2. Yang, J., Guo, Y., and Zheng, W. Handwritten text recognition using a hybrid CNN-LSTM model. Pattern Recognition Letters, 131, 123-130 (2020).
3. Liu, C., Zhang, Y., Lu, Y., and Xu, M. Handwritten text recognition using a hybrid CNN-GRU model. Pattern Recognition Letters, 125, 540-546 (2019).
4. Huang, X., Qiao, H., and Tang, Y. Handwritten text recognition using a hybrid CNN-GRU model with attention mechanism. Pattern Recognition Letters, 144, 50-56 (2021).
5. Cheng, L., Liu, C., Zhang, Y., and Lu, Y. A unified framework for online handwritten Chinese text recognition based on CNN and RNN. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51(6), 3298-3309 (2021).
6. Kang, W., Huang, Z., Liu, J., and Wang, J. Offline handwritten text recognition with hybrid CNN-LSTM model. Journal of Visual Communication and Image Representation, 66, 102767 (2020).
7. Su, W., Huang, Y., and Zhang, S. Handwritten text recognition based on hybrid CNN-LSTM model with attention mechanism. International Journal of Machine Learning and Cybernetics, 12(1), 183-193 (2021).
8. Cao, L., Wang, H., Wang, Y., and Wu, Y. A hybrid CNN-LSTM network for online handwritten Chinese character recognition. Frontiers in Psychology, 11, 622647 (2021).
9. Du, X., Zhang, Y., and Yang, Y. Handwritten Chinese text recognition with hybrid CNN-GRU model and channel attention. Frontiers in Psychology, 12, 770887 (2021).
10. Su, W., Huang, Y., & Zhang, S. (2021). Handwritten text recognition based on hybrid CNN-LSTM model with attention mechanism. *International Journal of Machine Learning and Cybernetics, 12*(1), 183-193.
11. Cao, L., Wang, H., Wang, Y., & Wu, Y. (2021). A hybrid CNN-LSTM network for online handwritten Chinese character recognition. *Frontiers in Psychology, 11*, 622647.
12. Du, X., Zhang, Y., & Yang, Y. (2021). Handwritten Chinese text recognition with hybrid CNN-GRU model and channel attention. *Frontiers in Psychology, 12*, 770887.
13. Mishra, A., Singh, V. K., & Saraswat, M. (2021). Hybrid CNN-LSTM Model for Handwritten Text Recognition in Devanagari Script. *Pattern Recognition and Image Analysis, 31*(4), 756-764.

14. Saba, T., Rehman, A., & Rashid, M. (2022). A Novel Hybrid Deep Learning Model for Offline Handwritten Text Recognition. *Journal of Imaging, 8*(2), 34. https://doi.org/10.3390/jimaging8020034
15. Khandare, A., & Kumar, P. (2021). Handwritten Document Recognition Using Hybrid Deep Learning Approach. *Neural Computing and Applications, 33*(13), 7127-7140. https://doi.org/10.1007/s00521-021-05713-7