Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

A Deep-Learning Technique for Converting Bengali Handwritten Answer-Book Images into Text

Moumita MOITRA, Malay Kumar MAJHI & Sujan Kumar SAHA*
Department of CSE, National Institute of Technology Durgapur, India
*sujan.kr.saha@gmail.com

Abstract: Computer-Assisted Assessment (CAA) involves the use of computer technology for educational assessment. Handwritten answer books are the primary medium for various levels of educational assessment in schools in India. Handwritten answer books pose the key challenge in adopting CAA in the Indian language, Bengali. These answer books need to be converted into machine-readable text before applying any tool for evaluation of the answers. The literature on Bengali handwritten word recognition primarily focused on considering the image of a single word as input. However, for automatic answer evaluation, the image of the whole page should be taken as input by the system. In this paper, we aim to develop a system for converting handwritten words from a student's answer book into machine-readable text. The proposed system follows a two-phase architecture. The first phase automatically segments the words using Mask Region-based Convolutional Neural Network (R-CNN) network. Then, the segmented words are used as input for the second phase of image-to-text conversion using a CNN-Transformer network. To train the models, we use an openly available dataset and an in-house dataset comprised of answer books collected from Bengali medium schools. The experimental results show that the proposed model is quite promising.

Keywords: Automatic Answer Grading, Handwritten Character Recognition, Word Segmentation, Deep Learning model.

1. Introduction

The use of computer technology in educational assessment is a widely explored research problem. Several systems have been developed to effectively use computers in educational assessment. Due to various advantages over the traditional manual evaluation, computer-assisted assessment (CAA) has been utilized in various education sectors. Numerous studies have been conducted that demonstrates the effectiveness of CAA. However, due to the lack of necessary infrastructure, use of CAA is hindered and the paper-based examination and assessment is still being continued for educational assessment in India. Most of the existing CAA frameworks needs the examination to be conducted on a digital platform and then automatic evaluation is done. These frameworks cannot be applied directly to paper answer books. In order to adopt CAA here, the first step is to convert the paper answer books into machine readable text.

Handwritten Text Recognition (HTR) is the process of converting handwritten script images into machine-readable text. Automatic Recognition of Handwritten Characters has been a widely explored research area for decades. In the early days, the focus was to convert the images of handwritten characters or digits into corresponding machine-readable characters, and the area was named Handwritten Character Recognition (HCR). With the advancement of information and communication technologies, new application areas emerged, and it became necessary to recognize handwritten words and text. Ample research effort has been devoted to developing HTR systems in various languages (AlKendi et al., 2024; Teslya and Mohammed, 2022; Moitra and Saha, 2024).

Bengali is the sixth most widely spoken languages in the world. It is the second most widely spoken language in India and the official language of Bangladesh. Bengali is the

primary medium of education in a large number of Bengali medium schools in West Bengal and Bangladesh. Handwritten text plays the most crucial role in content delivery and assessment there. So, recognition of handwritten Bengali text is crucial for smart education and assessment. Recognition of Bengali handwriting is comparatively more difficult than English due to its complex shape, matra, and the presence of numerous conjunct characters. Several systems have been developed for recognizing Bengali characters (Bhowmik et al., 2016; Chaudhuri and Bera, 2009; Dutta et al., 2021; Keserwani et al., 2019). Most of these systems were built on handwritten document images written by adults, where handwriting is clearer, matured and more structured. On the other hand, the handwriting of school students is more inconsistent, inter-word and intra-word spacing is not uniform, and contains overlapping of characters. Therefore, recognition of school students handwriting is more challenging. We did not find any openly available system that takes an image of a page of a handwritten Bengali answer book and accurately converts it into the corresponding text.

This paper focuses on developing a system for converting handwritten words from a student's answer book into machine-readable text. A two-phase architecture is proposed for the task. The first phase performs *word segmentation*. It basically marks the word boundaries from the input image of a handwritten whole page containing several lines and words. The output of the first phase, segmented word images, is fed into the second phase, which recognizes the words. The word segmentation model is developed using Mask Region-based Convolutional Neural Network (Mask RCNN). The core component of the RCNN architecture contains a Region Proposal Network (RPN) to generate candidate regions likely containing words, and a Fully Convolutional Network (FCN) is used for pixel-to-pixel segmentation. The word recognition model consists of a Convolutional Neural Network (CNN) followed by a positional encoding layer and an encoder module of the Transformer. The output from the encoder is then finally fed into a linear layer, which produces the predictions.

For the training of the system, we collected answer books from middle school students from multiple Bengali medium schools and extracted and labeled words from them. We evaluated the performance of the individual phases and the whole system using an in-house test dataset comprising images of school students' handwritten pages. We created baseline models in both phases using standard techniques to compare the performance of the proposed systems. The proposed technique outperformed the baseline models in both phases. The system achieved an overall accuracy of 78.12%. To evaluate the potential of the proposed technique, we also tested the system using a dataset containing images of Bengali handwritten pages collected from online sources. Then the system obtained an accuracy of 93.62%. A summary of the relevant literature, the details of the individual components of the proposed system, datasets used to train and test the system and the experimental results are discussed in the subsequent sections of the paper.

2. Literature Review

A number of researchers have proposed various methods for recognizing offline handwritten text in English and other languages. Some models have been developed for various Indian languages, including Bengali. In this section, we review some of these existing works.

2.1 Works on Handwritten Word Segmentation

The segmentation of handwritten text is a well-explored area in document analysis and optical character recognition (OCR). Yet, it presents significant challenges, particularly with complex scripts like Bengali.

In earlier techniques, methods such as Projection Profile (Chaudhuri and Bera, 2009), Smearing method (Roy et al., 2008), Water Reservoir Method (Pal et al., 2001), Connected Component Analysis (Khandelwal et al., 2009), Scale Space method (Manmatha and Rothfeder, 2005) were commonly used to segment characters, lines, and words from handwritten document images. However, all the above methods have some major weaknesses. Many of these methods were extensively explored in Indian handwritten text

segmentation, and some of them are extremely sensitive to parameters like slanted and skewed at the word level, irregular gaps, noise, and language-specific complexities. Some are extremely sensitive to parameter settings and are effective for only one or a few specific types of languages or handwriting.

Researchers also combined multiple techniques to develop hybrid approaches. For instance, Mullik et al. (2015) developed a hybrid approach for text line segmentation in handwritten Bangla documents. They started by smudging the document image to reduce white spaces between words, followed by segmenting the image at its whitest pixels. Thinning was then used to address multi-line components, identifying the most likely separation points. In recent work, Rakshit et al. (2023) used region-based filling operation and an innovative light projection technique that differentiates text lines from the background. However, we feel that the method heavily depends on low-level preprocessing, such as binarization and computation of connected components.

The most recent methods have shifted towards deep learning approaches, overcoming challenges like line skewness, uneven spacing, and irregular paragraph shapes seen in earlier techniques. CNN-based semantic methods (Keserwani et al., 2019; Inunganbi et al., 2021; Dutta et al., 2021; Pramanik and Bag, 2020) were broadly explored for text segmentation. Jindal and Ghosh (2023) utilized an instance segmentation method with an R-CNN, a region proposal network (RPN), and a region of interest pooling layer to segment ancient Indian handwritten scripts.

2.2 Works on Handwritten Word Recognition

The early works focused on recognizing isolated characters by using classical machine-learning techniques. For instance, Das et al. (2010) developed a Multi-Layer Perceptron (MLP) and SVM-based classifier for Bengali HCR. They obtained an average recognition rate of 79.25% using MLP and 80.51% using SVM. Basu et al. (2012) designed a classifier using an MLP for Bengali HCRs using a feature set of 76 elements. Their reported accuracy on the training dataset was 86.46%, and on the testing dataset, it was 75.05%. Roy et al. (2014) implemented a Hidden Markov Model (HMM) based architecture for offline Bengali word recognition. They used zone segmentation techniques to model the various zones and used HMM in the middle zone.

Later, the trend shifted towards the use of deep learning techniques. Purkaystha et al. (2017) developed a convolutional deep learning model that uses kernels and local receptive fields and then employs densely connected layers for the recognition task. They tested their system on the BanglaLekha-Isolated dataset and obtained a testing accuracy of 89.93%. Gupta et al. (2023) used the CNN-Transformer model and built handwritten OCR from eight Indian languages. They worked on word recognition from handwritten text. Haque et al. (2022) proposed an approach for Handwritten Bangla Character Recognition using CNN and the YOLOv5 algorithm, achieving 96% accuracy in character recognition and 91% in word recognition.

However, most of the works mentioned above used an image of a word as input to the system. In the current paper, we consider the image of a whole handwritten page that contains multiple text lines and words as input. So, those recognition methods are not directly applicable. Therefore, we proposed using a segmentation phase before the recognition.

3. Methodology

In this work, we present a two-phase technique for converting an image of a full handwritten page into corresponding machine-readable text. The workflow of the technique is summarized in Figure 1, and the working of the individual phases is discussed below.

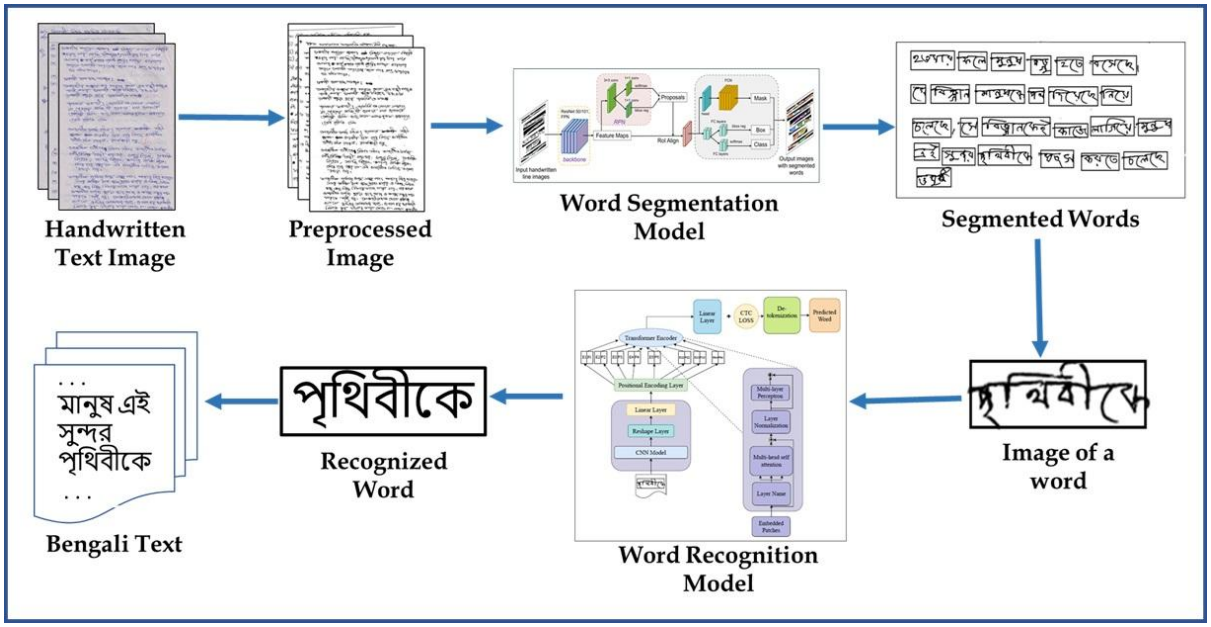

Figure 1. Workflow of the proposed system for converting handwritten page images into text.

3.1 Bengali Handwritten Word Segmentation

This phase aims to extract images of individual words from images containing several words. First, we developed the baseline system using a CNN-based semantic segmentation model. Then, we designed an advanced instance segmentation model using Mask R-CNN, which provided more refined segmentation capabilities.

3.1.1 Baseline System

A CNN-based encoder-decoder architecture for word segmentation has been used as the baseline. This task has been treated as a semantic segmentation problem with two pixel-level classes: words and background. The encoder extracts semantic features through convolutional and pooling layers, while the decoder upsamples and performs pixel-wise classification to generate a dense probability map. The model is trained on a pixel-level annotated dataset, with pre-processing applied to handle challenges like touching characters and marginal noise in handwritten scanned documents. The model uses 5×5 filters, stride 1, and max-pooling for downscaling, doubling the number of filters at each scale to effectively encode complex and fine-grained features. The final output is a dense pixel-wise probability map that predicts whether each pixel belongs to a word or the background. This ensures robust and accurate segmentation of words from scanned handwritten document images. This method aimed to predict the word boundaries within the images directly. However, during experiments, we found that the performance was suboptimal, particularly in handling documents with dense text and small handwriting, where the model struggled to identify word boundaries accurately.

3.1.2 Mask R-CNN based Segmentation

Due to the suboptimal performance of the CNN-based approach, particularly in handling documents with dense text and small handwriting, we employed an instance segmentation method using Mask R-CNN. This method integrates object detection with semantic segmentation by evolving previous methods, including RCNN, Fast RCNN, and Faster RCNN. As shown in Figure 2, Mask R-CNN architecture consists of three key stages: (a) generation

of region proposals, (b) classification of each proposal, and (c) segmentation of each proposal to predict a pixel-level mask for the detected objects.

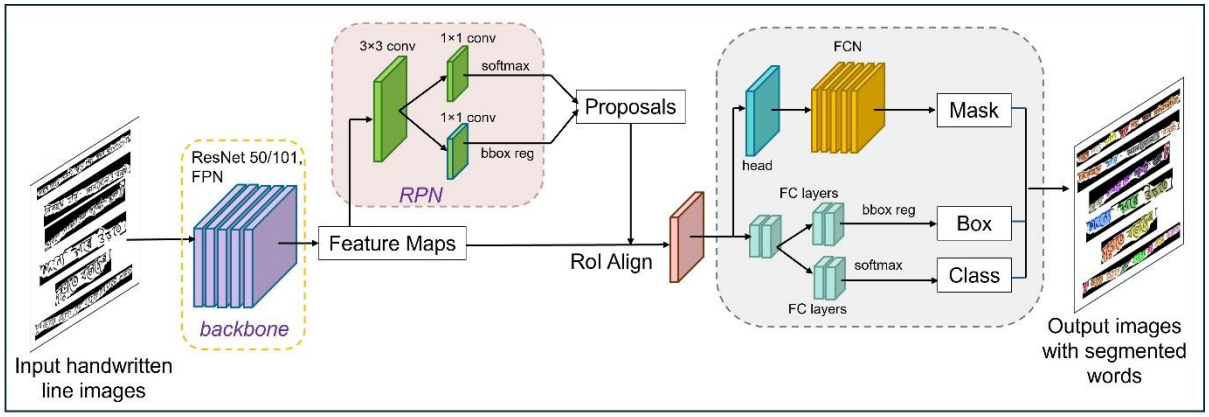

Figure 2. The Mask R-CNN architecture for word segmentation.

Here, we employed full-page handwritten answer sheet images for segmenting or extracting words. A fully convolutional network (FCN) is used for pixel-to-pixel segmentation in the Mask R-CNN, which also comprises the Faster R-CNN for object discovery. We can see a backbone network responsible for convolutions and the generation of feature maps. The backbone network in Mask R-CNN is typically a pre-trained convolutional neural network, such as ResNet, ResNeXt, VGGNet, etc. The FPN is added to this backbone network to create a feature pyramid. ResNet layers act as feature extractors, and the FPN helps to capture features at multiple scales, which is particularly useful for detecting words of varying sizes in handwritten documents. This technique employs a Region Proposal Network (RPN) to generate candidate regions likely containing words, predicting objectness scores and bounding box coordinates. The ROI Align layer ensures precise feature map alignment for each proposal, correcting quantization errors from traditional ROI pooling. This improves word detection and segmentation by retaining spatial feature alignment, which is essential for handwritten documents.

3.2 Bengali Handwritten Word Recognition

This phase aims to convert images of individual words into the corresponding text.

3.2.1 Baseline Model

This study first used a CNN-based architecture with CTC loss as the baseline model for handwritten Bengali word recognition. The CNN acts as a feature extractor, capturing spatial patterns through convolutional and pooling layers. Then, a BiLSTM layer processes the extracted features to model sequential dependencies in the text. The output is passed through a Softmax layer, producing character probabilities, and the final word is obtained using Connectionist Temporal Classification (CTC) loss, which aligns predictions with the target sequence without requiring explicit segmentation. The CNN consists of four convolutional layers, each followed by max-pooling to progressively extract hierarchical representations. The BiLSTM comprises two bidirectional layers with 256 units each, ensuring the model captures both forward and backward contextual dependencies in handwritten words. The model is trained with the Adam optimizer and a learning rate of 1e-4, optimizing character-

level predictions for improved recognition accuracy.

3.2.2 A CNN-Transformer Model for Recognition

In the experiments using the CNN-BiLSTM models, it was found that both models failed to correctly predict words containing constant conjuncts and diacritic forms of vowels (Juktakhors). These limitations motivated us to explore further. It was felt that the mapping of images of handwritten characters to corresponding characters of the Bengali vocabulary could be interpreted as a sequence-to-sequence mapping problem. Transformers were proven to be more efficient in sequence-to-sequence mapping tasks since they allow parallel processing of tokens and employ self-attention mechanisms. So, the final model is based on *CNN-Transformer* architecture and is shown in Figure 3.

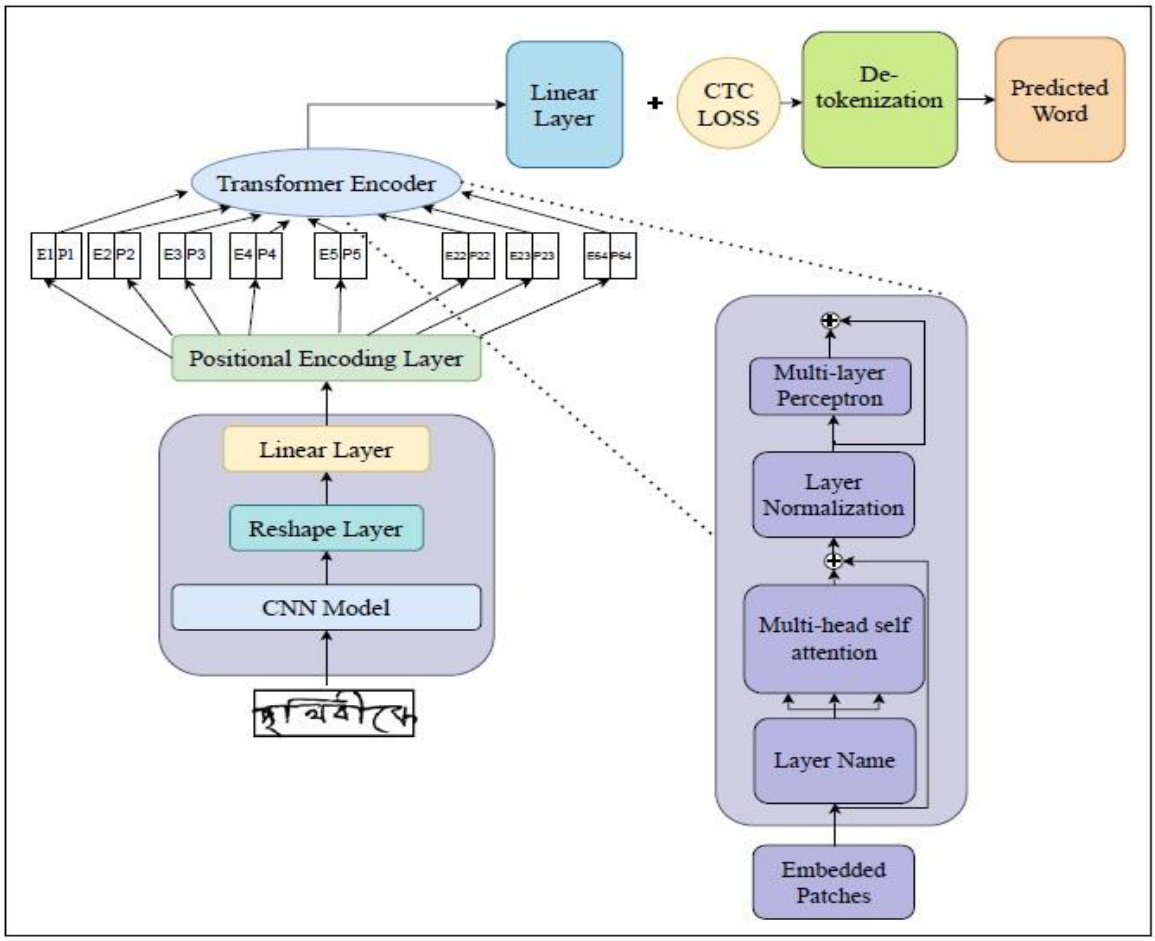

Figure 3. The workflow of the proposed CNN-Transformer model

The proposed CNN-Transformer model has thirteen convolution layers, and the last layer of the CNN is a max pool layer of pool size (2, 2). The depth of CNN was increased by stacking multiple convolutional layers, each with a non-decreasing number of filters (e.g., 64, 128, 256, 512, etc.). This design choice enables the network to learn a diverse range of features at different levels of abstraction, enhancing its capability to represent intricate patterns in image data effectively. Each convolution layer internally uses an activation function, which adds an element of non-linearity and helps the model learn complex patterns and training until convergence. Here, ReLU is used. Additionally, batch normalization normalizes the activation within each mini-batch, leading to faster training and improved model stability by reducing internal covariate shifts and acting as a regularizer. The max pool layer helps reduce computational complexity and control overfitting by extracting dominant features while

retaining crucial spatial information. The hyperparameters of the model were carefully tuned to achieve optimal performance. After processing through the CNN, the output feature map is reshaped and permuted to prepare it for linear transformation. The reshaping operation converts the feature map into a suitable format for subsequent processing by the transformer's encoder layers. For every batch of feature maps that the CNN generates, a positional encoding is applied to inject positional information into the sequence representation before passing it through the encoder layers.

Finally, the transformer's output sequence representation is passed through a linear layer to predict the target sequence. This linear layer maps the transformer's output from the model's internal representation to the target vocabulary size, generating the final predictions. Lastly, the de-tokenization module is used to convert the output tensor from our transformer into the corresponding predicted word.

4. Experimental Results

This section discusses the results obtained from the experiments.

4.1 Dataset for Word Segmentation

There are certain openly available datasets; however, the handwriting of school students contains specific challenges. The openly available Bengali handwritten datasets do not cover these domain-specific issues and are not very suitable for our development. Therefore, we created a domain-specific dataset. To create the dataset, we collected student answer books of grade V-VII students from Bengali-medium schools. These answer books are scanned, and full-page images are used to prepare the dataset.

We first cleaned the images to remove the noises that could interfere with segmentation accuracy. Then, the images were converted into binary format. This binary conversion is crucial as it reduces computational complexity and focuses the model's attention on the text structure, facilitating more effective learning and analysis. Next, the images are manually annotated with word regions using VGG Image Annotator (VIA) to prepare the training data. During annotation, the individual units (essentially words) were marked on a zoomed image by a set of points forming a closed polygon. The training data comprised 205 annotated handwritten answer sheet images. The test set consists of 20 randomly selected answer sheet images, comprising 2098 words in total.

4.2 Dataset for Word Recognition

We used a deep learning architecture to develop our model, so a large training dataset is required. An openly available Bengali handwritten dataset, BN-HTRd[1], is used for the experiments. The BN-HTRd dataset (Rahman et al., 2023) is derived from the BBC Bangla News corpus and encompasses 788 full-page images meticulously compiled from the contributions of 150 distinct writers. Furthermore, it comprises a total of 108,147 instances of handwritten words, organized across 13,867 lines and incorporating 23,115 unique words.

4.3 Evaluation Metrics

To evaluate the performance of the segmentation models, we used % accuracy. The accuracy is computed as the ratio of correctly segmented words to the total number of words in the test data.

The primary metrics used for the evaluation of word recognition models are the Word Recognition Rate (WRR) and Character Recognition Rate (CRR). WRR measures the percentage of correctly recognized words in the dataset, providing an overall assessment of

[1] https://data.mendeley.com/datasets/743k6dm543/1

the system's performance at the word level. On the other hand, CRR evaluates the percentage of correctly recognized characters, presenting a fine observation capability of the model to correctly predict individual characters within word.

4.4 Results of the Word Segmentation Model

In this study, we extracted words from full-page handwritten documents using a baseline semantic segmentation approach employing a CNN model using the aforementioned dataset. The model was trained for ten epochs, each comprising 1,000 iterations. The model achieved an accuracy of 73.02% on the in-house school dataset. We have also tested the system using two other test datasets. On the Bengali online handwritten test images, this model correctly detected 1367 words among 1779 total words with accuracy of 76.84%, as shown in Table 1.

Next, we have adopted an instance segmentation approach for segmenting words from full-page handwritten images using the Mask R-CNN model. When we computed the word count-based segmentation accuracy, we found that the R-CNN model achieves 79.64% accuracy. However, it performed well when the model was tested using the online test dataset. For the Bengali online dataset, the model correctly detected 1701 words out of 1779 total words, achieving an accuracy of 95.61%. As the inter-line and inter-word gap is mostly uniform and the datasets mainly comprise matured handwriting, it identified the boundaries correctly.

Table 1. *Performance of Word Segmentation.*

Model	Dataset	Total Words	Detected Words	Accuracy
Baseline	In-house: answer books	2098	1532	73.02%
	Online Bengali	1779	1367	76.84%
Mask R-CNN	In-house: answer books	2098	1671	73.02%
	Online Bengali	1779	1701	95.61%

4.5 Results of the Word Recognition Model

The baseline model was trained for 120 epochs. The model achieved 81.34% WRR and 89.07% CRR when tested on an in-house dataset. We have also tested the system when it was trained using the BN-HTRd dataset. Then the WRR and CRR values are 73.4% and 82.62%, respectively. These values prove that the handwriting of school students is more difficult to recognize, and available open datasets alone cannot provide high accuracy. Furthermore, experiments were conducted using the proposed CNN-Transformer model. The transformer-based model showed a substantial performance improvement over the baseline. The model was trained until convergence, at 43 epochs. Then, it achieved a WRR of 91.74% and a CRR of 95.18%. The values are summarized in Table 2. This improvement proves the superiority of the proposed transformer-based model in the handwritten word recognition task.

Table 2. *Performance of Word Recognition.*

Model	Training Data	Test Data	WRR	CRR
Baseline	In-house	In-house	81.34%	89.07%
	BN-HTRd	In-house	73.40%	82.60%
CNN-Transformer	In-house	In-house	91.74%	95.18%
	BN-HTRd	In-house	84.81%	89.70%

4.6 Performance of the 2-phase Model

To assess the model's overall performance, we take a few images of handwritten pages from the in-house school dataset as input to the system, run both phases and compute the accuracy of the final word recognition. The performance is measured using percentage accuracy: how many words are correctly recognized by the final system divided by the total number of words. There we found that the system achieved an accuracy of 78.12%. When we tested the performance of the final system using online handwritten images, the system achieved an accuracy of 93.62%. The performance is relatively lower in the handwritten answer-book images because of higher errors in segmentation. When we analyzed the outputs, we found that several words were partially segmented and degraded overall performance.

5. Discussion and Future Work

We proposed a two-phase architecture for converting the image of a handwritten page into machine-readable text. The first phase performs word segmentation, and the second phase recognizes the segmented words. Deep-learning architecture is used in both phases. The system aimed to recognize handwritten answer books of middle-school students.

We found that the model could not segment the words when the words are densely written, and the inter-word gap is not uniform. The model tends to club multiple words into one segment. Again, in some cases, one word is segmented into multiple units due to a larger gap between individual characters of a word. The use of better deep learning architecture and language model-based post-processing might improve the performance of the system. Creation of larger domain-specific training data for segmentation and recognition can be another direction to work in the future.

Funding

The work has been partially funded by ANRF/SERB, India with Project File No: EEQ/2021/000687.

References

AlKendi, W., Gechter, F., Heyberger, L., Guyeux, C.: Advancements and challenges in handwritten text recognition: A comprehensive survey. Journal of Imaging 10(1), 18 (2024).

Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., and Basu D. K.: Handwritten Bangla Alphabet Recognition using an MLP Based Classifier. arXiv: arXiv.1203.0882 (2012)

Bhowmik, T.K., Parui, S.K., Roy, U., Schomaker, L.: Bangla handwritten character segmentation using structural features: a supervised and bootstrapping approach. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 15(4), 1–26 (2016).

Chaudhuri, B.B., Bera, S.: Handwritten text line identification in Indian scripts. In: 2009 10th International Conference on Document Analysis and Recognition. pp. 636–640. IEEE (2009)

Das, N., Das, B., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M.: Handwritten Bangla basic and compound character recognition using MLP and SVM classifier, arXiv:1002.4040. (2010).

Dutta, A., Garai, A., Biswas, S., Das, A.K.: Segmentation of text lines using multi-scale CNN from warped printed and handwritten document images. International Journal on Document Analysis and Recognition (IJDAR) 24(4), 299–313 (2021).

Gupta, M.K., Vikram, S., Dhawan, S., Kumar, A.: Handwritten OCR for word in Indic language using deep networks. In: 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN). pp. 389–394 (2023).

Haque, P., Salma, U., Chowdhury, R.: Bangla handwritten character and words recognition-based on the yolov5 algorithm. In: 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE). pp. 24–29 (2022).

Hossain, M.I., Rakib, M., Mollah, S., Rahman, F., Mohammed, N.: Lila-Boti: Leveraging isolated letter accumulations by ordering teacher insights for Bangla handwriting recognition. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 1770–1776. IEEE (2022).
Inunganbi, S., Choudhary, P., Manglem, K.: Meitei Mayek handwritten dataset: compilation, segmentation, and character recognition. The Visual Computer 37(2), 291–305 (2021).
Jindal, A., Ghosh, R.: Word and character segmentation in ancient handwritten documents in Devanagari and Maithili scripts using horizontal zoning. Expert Systems with Applications 225, 120127 (2023).
Keserwani, P., Ali, T., Roy, P.P.: Handwritten Bangla character and numeral recognition using convolutional neural network for low-memory GPU. International Journal of Machine Learning and Cybernetics 10, 3485–3497 (2019).
Khandelwal, A., Choudhury, P., Sarkar, R., Basu, S., Nasipuri, M., Das, N.: Text line segmentation for unconstrained handwritten document images using neighborhood connected component analysis. In: Pattern Recognition and Machine Intelligence: Third International Conference, PReMI 2009 New Delhi, India, December 16-20, 2009 Proceedings 3. pp. 369–374. Springer (2009)
Manmatha, R., Rothfeder, J.L.: A scale space approach for automatically segmenting words from historical handwritten documents. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), 1212–1225 (2005).
Moitra, M., Saha, S.K.: A review on handwritten text segmentation in Indian languages. *Int. J. Mach. Learn. & Cyber.* (2024). https://doi.org/10.1007/s13042-024-02448-1
Mullick, K., Banerjee, S., Bhattacharya, U.: An efficient line segmentation approach for handwritten Bangla document image. In: 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR). pp. 1–6. IEEE (2015)
Pal, U., Belaïd, A., Choisy, C.: Water Reservoir based approach for touching numeral segmentation. In: Proceedings of Sixth International Conference on Document Analysis and Recognition. pp. 892–896. IEEE (2001)
Pramanik, R., Bag, S.: Segmentation-based recognition system for handwritten Bangla and Devanagari words using conventional classification and transfer learning. IET Image Processing 14(5), 959–972 (2020).
Purkaystha, B., Datta, T., Islam, M.S.: Bengali handwritten character recognition using deep convolutional neural network. In: 2017 20th International Conference of Computer and Information Technology (ICCIT). pp. 1–5 (2017).
Rahman, M.A., Tabassum, N., Paul, M., Pal, R., Islam, M.K.: BN-HTRd: A benchmark dataset for document level offline Bangla handwritten text recognition (HTR) and line segmentation. In: Computer Vision and Image Analysis for Industry 4.0, pp. 1–16 (2023).
Rakshit, P., Halder, C., Sk, M.O., Roy, K.: A generalized line segmentation method for multi-script handwritten text documents. Expert Systems with Applications 212, 118498 (2023).
Roy, P.P., Dey, P., Roy, S., Pal, U., Kimura, F.: A novel approach of Bangla handwritten text recognition using HMM. In: 14th International Conference on Frontiers in Handwriting Recognition. pp. 661–666 (2014).
Roy, P.P., Pal, U., Lladós, J.: Morphology based handwritten line segmentation using foreground and background information. In: International conference on frontiers in handwriting recognition. pp. 241–246 (2008).
Teslya, N., Mohammed, S.: Deep learning for handwriting text recognition: existing approaches and challenges. In: 2022 31st Conference of Open Innovations Association (FRUCT). pp. 339–346. IEEE (2022)