Handwritten Text Segmentation Using U-Net and Shuffled Frog-Leaping Algorithm with Scale Space Technique

Moumita Moitra and **Sujan Kumar Saha***
National Institute of Technology Durgapur
mm.23cs1108@phd.nitdgp.ac.in, sksaha.cse@nitdgp.ac.in

Abstract

The paper introduces a new method for segmenting words from handwritten Bangla documents. We found that the available handwritten character recognition (HCR) systems do not provide the desired accuracy in recognizing the text written by school students. Recognizing students' handwritten text becomes challenging due to certain factors, including a non-uniform gap between lines and words, and ambiguous, overlapping characters. The performance may be improved if the words in the text are segmented correctly before recognition. For the segmentation, we propose a combination of U-Net and a modified Scale Space method enhanced by the Shuffled Frog-Leaping Algorithm (SFLA). We employ the U-Net model for line segmentation; it effectively handles the variable spacing and skewed lines. After line segmentation, for segmenting the words, we use SFLA with Scale Space, allowing adaptive scaling and optimized parameter tuning. The proposed technique has been tested on two datasets: the openly available BN-HTR dataset and an in-house dataset prepared by collecting Bengali handwritten answer books from schools. In our experiments, we found that the proposed technique achieved promising performance on both datasets.

1 Introduction

The process of converting handwritten text into machine-readable form is known as handwritten character recognition (HCR). When the input to the HCR system is an image of a whole handwritten page containing several lines of text, it is essential to extract the individual words for better processing. Handwritten word segmentation is the process of extracting the words from a handwritten text image. Although segmentation primarily focuses on separating words, it also refers to separating lines or characters in some domains.

In the case of printed documents, the space between any two consecutive words, characters, and lines almost always remains consistent. However, in handwritten documents, the space varies since each person's writing style is unique. When addressing handwritten word segmentation, intricacies arise due to the presence of slant and skew at the word level and the non-uniform inter-word, intra-word, and inter-line gaps. Also, noise and other language-specific features like matra, zone, and compound characters make the task more challenging. The task's difficulty increases when the skewness, non-uniformity in gaps, and noise level increase. For instance, we observed that word segmentation from Bengali handwritten school answer books is more challenging than many other domains.

In our study, we have developed a hybrid method for segmenting handwritten words. First, we employ the U-Net model (Delil et al., 2021; Xiao et al., 2022) for segmenting lines from handwritten documents. The U-Net model is based on the convolutional network architecture renowned for its success in biomedical image segmentation. The U-Net model is particularly effective in this task due to its convolutional layers that adeptly learn handwriting characteristics such as stroke width and line curvature. Its ability to process images at multiple resolutions allows it to adeptly segment text lines that are skewed or densely packed, overcoming typical challenges in handwriting recognition.

Following line segmentation, we utilize the Shuffled Frog-Leaping Algorithm (SFLA) (Eusuff et al., 2006) to enhance the performance of the traditional Scale Space method (Manmatha and Srimal, 1999), previously applied to handwritten word segmentation in English and other languages. The efficacy of the Scale Space method is highly

dependent on the precise selection of scaling parameters. The integration of SFLA automates this parameter tuning, aiming for an optimal scaling that improves segmentation. Consequently, this enhances the segmentation accuracy and overall performance of the Scale Space method.

To compare the performance of the proposed technique, we have also considered three other widely used methods: the vanilla Scale Space method (Manmatha and Srimal, 1999), the Connected Component Analysis method (Khandelwal et al., 2009), and the Stroke Width Analysis method (Shivakumara et al., 2022). All four methods are implemented and tested with the same dataset to measure their relative performance. The test dataset contains images from two different sources: an open online Bengali HCR dataset - BN-HTR (Rahman et al., 2023), and our in-house dataset that contains images of handwritten answer books of grade six Bengali medium students. In our experiments, we found that the proposed technique performs better than others. By integrating the Shuffled Frog-Leaping Algorithm with the Scale Space method (SFLA-SS), we attained an accuracy of 93.64% on the BN-HTR dataset and 56.84% on our in-house dataset. The subsequent application of the U-Net model in conjunction with SFLA-SS resulted in an enhanced accuracy of 96.18% on the BN-HTR dataset and 69.29% on the in-house dataset.

2 Related Work:

Bengali, also known as Bangla, is one of the major languages in the Indian subcontinent. Numerous research works have already been published on handwritten Bengali character recognition. It's the seventh most spoken native language in the world, with nearly 230 million total speakers. The Bengali script, also known as Bangla, is an abugida, which means each character represents a consonant with an inherent vowel, and other vowels are represented with diacritics. It also includes a number of compound characters and conjuncts that combine two or more characters. This complexity makes it difficult to distinguish where one word ends and another begins. Many research works are investigated for handwriting text recognition and handwritten document segmentation of different Indian scripts Jindal and Ghosh (2023); Priyadarshi and Saha (2021); Inunganbi et al. (2021). Some of the popular text segmentation methods are the Projection Profile method, Connected Component Analysis (CCA), Scale Space method, Stroke Width Analysis, and machine learning and deep learning-based methods. There are some studies in the literature that focus solely on segmenting handwritten Bengali lines and words. Pal et al. (Pal and Datta, 2005) applied the Projection Profile method for text line segmentation from Bengali handwritten documents. MamathaH et al. (Mamatha and Srikantamurthy, 2012) Proposed a segmentation technique for extracting individual text-lines from handwritten document images using morphological operation and run-length smearing algorithm (RLSA). A word in Bengali script can be divided into three horizontally adjacent zones called the bottom zone, middle zone, and upper zone. These zones were utilized in this article (Khandelwal et al., 2009) to identify words that were contained within a single line. Pramanik et al. (Pramanik and Bag, 2020) proposes a method for recognizing handwritten Bengali and Devanagari words. It detects and corrects skew within words, estimates the headline, and segments words into meaningful pseudo characters. Capobianco et al. (2021) used the scale space approach for word segmentation from handwritten historical documents. Manmatha and Rothfeder (2005) proposed a tri-level handwritten text segmentation technique where they have used the scale space method for word-level segmentation of handwritten documents. To our knowledge, this is the only work that recognizes Bengali words from images. All these methods have some drawbacks in the case of oblique texts and critical overlapping situations. Also, most of these methods were applied on datasets where the space between lines and words was clear enough and noticeable. After observing the efficacy of all existing methods, we discerned that there is still significant scope for the enhancement of the word segmentation process from handwritten document images.

3 Existing Algorithms for Segmentation

As we found in the literature, several techniques have been applied to segment handwritten words. These can be broadly grouped into learning-based and non-learning-based techniques. Learning-based techniques refer to those using a Machine learning or deep learning technique applied to training data to generate the model. On the other hand, non-learning-based techniques use an algorithm to

analyze the pixels, their gaps, and other properties to perform the segmentation. As the training data is unavailable in our domain of interest (Bengali handwritten answer books from schools), we consider the non-learning-based techniques in this study. As per our study of our domain of interest and theoretical understanding, we have chosen three methods that we implemented first. These are discussed below.

3.1 Connected component analysis (CCA)

Connected component analysis is a popular technique that finds a wide range of applications in computer vision and pattern recognition. CCA works by grouping connected pixels with the same pixel intensity value. Based on the connectivity, CCA divides the image into individual components. In the handwritten text, CCA can help determine separate words, characters, or even parts of characters by identifying these connected components. In the case of handwritten text, this approach can be practical if the units (characters or words) are well-separated and do not overlap. This primarily works well when the writing is cursive, and a word is written in a single stroke. However, it might not work well in overlapping handwriting where individual letters and words are not neatly separated. Therefore, the performance of CCA might be influenced by factors like writing style, ink spread, or noise in the image.

3.2 Stroke width analysis (SWA)

This method begins by estimating the Stroke Width (SW) through the detection of sequences of continuous black pixels in every row of the image. Once these sequences are identified, their lengths are measured, through which a representative value for the stroke width is obtained. Such stroke width acts as a criterion for executing vertical and horizontal smoothing processes to identify word-level components. In vertical smoothing, the image is traversed vertically: if a sequence of black pixels is smaller than or equal to the stroke width (SW), they are considered noise. Those pixels were replaced by white pixels, which represent the background of the image. Otherwise, the pixels are considered to be part of the foreground. In horizontal smoothing, the image is scanned from left to right: if a series of white pixels is shorter than five times the stroke width (SW), they're viewed as part of the text. Otherwise, these pixels are seen as noise. After the above two steps, a morphological opening operation is applied to merge the characters of a word into a single component. After applying the smoothing and morphological operations, the image is divided into word-level components Shivakumara et al. (2022); each should ideally represent a distinct word.

3.3 Scale Space Method

The Scale-Space technique is another algorithm for word segmentation, particularly beneficial in handwritten texts. The scale-space theory focuses on the principle that scale is a critical factor in any physical observation, meaning that objects and features have significance only at specific scales. The concept applies a convolution of a function f(x, y) with a two-dimensional Gaussian kernel to generate a sequence of smoothed signals L(x, y; t). These signals, forming a linear scale-space representation, offer a series of progressively blurred versions of the original two-dimensional image, thus revealing information at various scales.

$$G(x, y; t) = \frac{1}{2\pi t} e^{-\frac{x^2+y^2}{2t}} \tag{1}$$

In the above equation, the scaling parameter *t* indicates the variance for the Gaussian filter at the scale level where t $\geq$ 0.

$$L(x, y; t) = G(x, y; t) * f(x, y) \tag{2}$$

In Scale-Space representation, an image is transformed across various scales by a convolution operation, as shown in Equation 2. Here, L(x, y; t) denotes the scale-space image, 'f(x, y)' is the original image, and '*' indicates the convolution operation. As we increase the scale level 't,' it progressively reduces the image's fine details as a result of the smoothing effect of the Gaussian kernel. This phenomenon is encapsulated in Equation 2, where the convolution operation (f * G) modifies the shape of the image, smoothing it across the increasing scale levels (Capobianco et al., 2021). Subsequently, specific features, such as lines, curves, or other potential indicators of word boundaries, are extracted at each scale. Scaling level, Gaussian filter size, and sigma are the main components to segregate Bengali handwritten words from an image. These features are then analyzed to discern potential word boundaries to identify the scale where these boundaries are the most distinct. Selecting the correct scale is the most crucial part of this method, which requires careful tuning.

4 Proposed Method

Here, we discuss our proposed technique for word segmentation from a handwritten document image. Initially, we implemented the three algorithms mentioned in Section 3 for word segmentation. There, we observed that those often faced difficulty in segmenting the lines accurately, especially in our in-house school dataset. This is because of the high non-uniformity in inter-line gaps in the images. Then, we decided to employ U-Net, which can work well with less training data. The synergy between the U-Net model and the SFLA provides a robust framework for handwritten text segmentation. The U-Net's powerful line extraction capabilities, coupled with the SFLA's efficient word boundary optimization, result in a segmentation process that is both precise and adaptable to the complexities of handwritten documents. So, the final model we proposed combines the U-Net model and the shuffled frog leaping algorithm (SFLA) algorithm along with the Scale-Space method. The workflow is summarized in Figure 1 and is discussed below in detail. As the input images contain various noises, we first applied a few preprocessing steps, like image binarization and noise reduction, to clean the input image; this preprocessing is not shown in the figure.

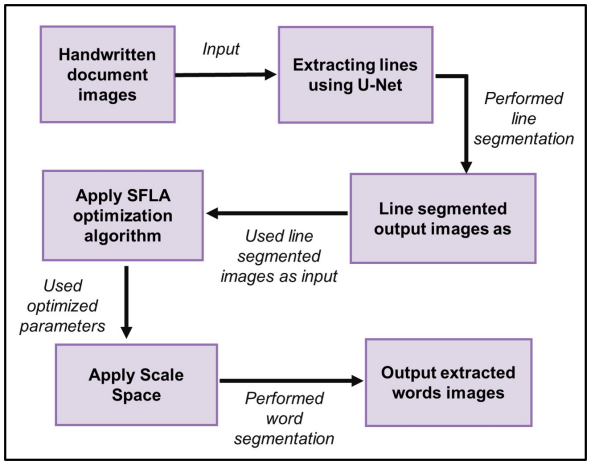

Figure 1: Workflow of the proposed method.

4.1 U-Net model

U-Net is a popular deep-learning architecture for semantic segmentation. U-Net model is basically a U-shaped convolutional neural network that was first used in the field of medical image segmentation. It is specifically designed for image segmentation, aiming to assign a label to each pixel in an image (Dong et al., 2021). U-Net model is a specific symmetric instance of the encoder-decoder network structure, with skip connections from layers in the encoder to the corresponding layers in the decoder. The specialty of this architecture is to perform well even with limited training data. Its structure is composed of an input layer, a contracting (encoder) path, a bottleneck, an expansive (decoder) path, and a bottleneck layer (latent layer or bridge layer) acting as a bridge connection. At the input layer, the model accepts 512x512 pixel images in grayscale. The output layer concludes the process with a 3x3 convolution to narrow down the output channels, followed by a 1x1 convolution with a sigmoid activation to predict pixel class probabilities. The model is optimized with the Adam algorithm, using a learning rate 1e-4, and trained against a binary cross entropy loss function, which is appropriate for binary classification.

4.2 Scale Space Method with SFLA (SFLA-SS)

Now, we discuss tuning the hyper-parameters of the scale-space technique using SFLA for handwritten Bengali word segmentation. While applying the vanilla Scale Space method we have chosen parameters through tuning of the experimental process. We conducted a series of experiments where different scale values were applied and the performance was measure based on its accuracy in segmenting the handwritten words. In addition to the tuning method, we also followed the existing literature on scale-space theory (Manmatha and Srimal, 1999), which provides a theoretical basis for understanding how different scales might affect the segmentation process.

In our study SFLA employs population-based heuristic search techniques and is typically used to solve convoluted optimization concerns. This meta-heuristic approach, which congregates the Memetic Algorithm (MA) and Particle Swarm Optimization (PSO) method, was first discoursed by Yusuf and Lansey in 2003. The shuffling meta-heuristic optimization method uses a population of conceivable solutions to facilitate local particle search and global information (Eusuff et al., 2006) conveying amongst groups in the initial population of P randomly produced frogs in D-dimensional space. Each frog is regarded as a particle in SFLA. The frogs are ranked in ascending or descending order using an objective

function (Han et al., 2013). The M/N number of frogs is distributed to each memeplex if there are M number of frogs and N number of memeplexes. According to this, the foremost memeplex has the frog with the highest fitness value demarcated by an objective function, and the second memeplex possesses the next-best frog. As long as M^{th} frog does not obtain the M^{th} memeplex, this process will continue. Iteration-wise, following this order, $(M+1)^{th}$ frog with the loftiest fitness value is again allocated to the first memeplex. Once the population search mechanism is over, the best frog is addressed as X_{NBest} whereas the worst one is addressed as X_{NWorst}. In the entire population, the frog with the global best fitness value is opted as the global best frog and is identified as X_G. In every memeplex, the worst frog is moving in the direction of the best or ideal one. If it is determined after each iteration that the new position is superior to the old one, the position is improved. The movement of X_{NWorst} has been obtained by Equations 3 to 5.

$$[M^T]_N = rand(X_{NBest} - X_{NWorst}) \quad (3)$$

In Equation 3, $[M^T]_N$ identifies the movement of the frogs with T^{th} iterations within the N memeplex with worst fitness value. Hence, rand() is a random number generation method that generates the numbers between 0 and 1.

$$M_{min} \leq [M^T]_N \leq M_{max} \quad (4)$$

From Equation 4, M_{min} and M_{max} both determine that the maximum number of alterations are permitted for the worst frog in the memeplex.

$$tX_{new} = X_{NWorst} + [M^T]_N \quad (5)$$

In Equation 5, X_{NWorst} denote the positions of the worst. If a better solution X_{new} is produced, then it replaces the X_{NWorst}. Accordingly, if no solution is attained, a new solution is picked randomly to replace X_{NWorst}. The memeplex is then jumbled together to share information and reallocated for the subsequent search process after being refreshed and reordered within local search durations.

This population-based meta-heuristic search method is utilized to tune the hyperparameters of the Scale-Space technique for word segmentation. To dissect an image's content at various granularities, scale-space space theory, a concept in image processing and computer vision, calls for portraying the image at several scales. Scale-space carries its cue from the human optic system, which is capable of perceiving objects and patterns in a range of scales, from minute fragments to comprehensive structures. A convolution of f(x, y) with the two-dimensional Gaussian kernel is used to yield a lineage of emanated signals L(x, y; t) for the linear scale space miniature of consecutive signals with arbitrarily sized signals for a two-dimensional image (Lee, 1983).

5 Experimental Setup

5.1 Dataset Collection

As we could not find any work or dataset on handwritten character recognition from Bengali student answer books, there needs to be a standardized dataset composed of student answer sheets. In the absence of a suitable dataset, we collected some handwritten Bengali images to test the proposed system. We have collected a few images from the open online BN-HTR dataset (Rahman et al., 2023), and some images are taken from an in-house dataset. The in-house dataset is collected from a Bengali medium school where the images of answer books of students in grades six and seven were scanned using a mobile scanner. The images of the whole pages were used as input to our system. Although the BN-HTR dataset images are cleaner, the images in the in-house dataset introduced an array of challenges that added complexity to the segmentation task, including issues like skew, curves, and closely spaced or touching lines of text; unequal intra-word and inter-word space; noise including poor paper quality, and scanning distortion. The test dataset contains eight images, four of which are taken from the in-house dataset, and the remaining four images are taken from the BN-HTR dataset. Figure 2 shows a few representative images from both sources.

5.2 System Evaluation Measure

In the process of performing word segmentation on handwritten documents, several problems can arise due to the complex handwriting style. Segmentation of words from handwritten Bengali documents presents unique challenges due to the specific structural and orthographic features of the scripts. The presence of compound characters, Matra, and Di-

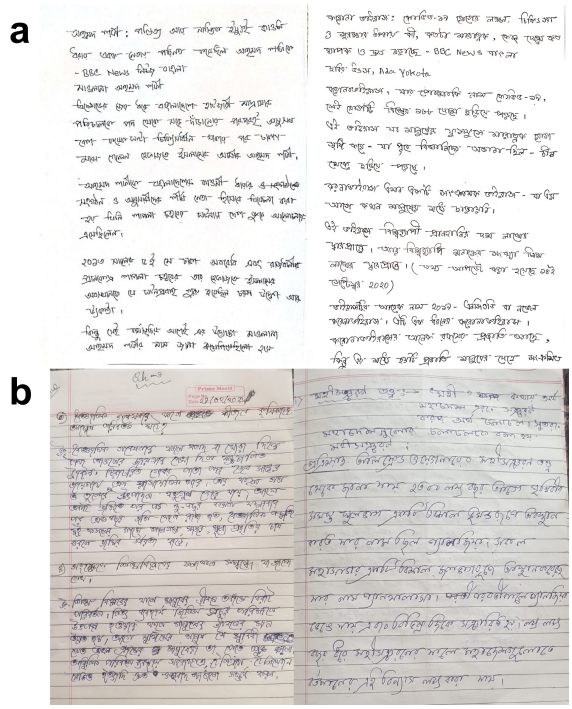

Figure 2: Representative images from the dataset used (a) BN-HTR dataset, (b) handwritten images collected from the school.

acritical marks creates a significant challenge for segmenting words from handwritten Bengali text. In our study, we have identified each word within a bounding box.

When we mark a machine-generated segmentation as correct or not, we might have considered two classes - correct and wrong. However, when we studied various output files, we found incorrect segmentation can be categorized into three classes: over-segmentation, under-segmentation, and partial-segmentation. And perfect-segmentation is the correctly segmented class where a single bounding box identifies the whole word. In the case of over-segmentation, a single word is inaccurately identified in multiple parts. This may occur often in Bengali text segmentation due to inconsistencies in handwriting, such as variations in space between letters within a word or Matra connecting some part and the rest not. On the other hand, under-segmentation is a situation where more than one word is mistakenly identified as a single word. This happens typically due to an unequal gap between words. Another problem occurred: partial segmentation, where a word is segmented, but some parts are incorrectly left out or included in another word.

6 Result and Discussion

To assess the effectiveness of the proposed methods, we implemented some existing techniques and tested them using both datasets. The existing techniques are Connected Component Analysis, Stroke Width Analysis, and Scale space method. Then, we implemented SFLA-SS for both line and word segmentation. The segmentation process was executed on the test set images, followed by a detailed evaluation based on the factors outlined earlier.

6.1 Performance of SFLA-SS

In Figure 3, we have shown the outputs of the four different methods applied to an in-house image containing student answers. The figure shows that SFLA-SS demonstrated the best performance.

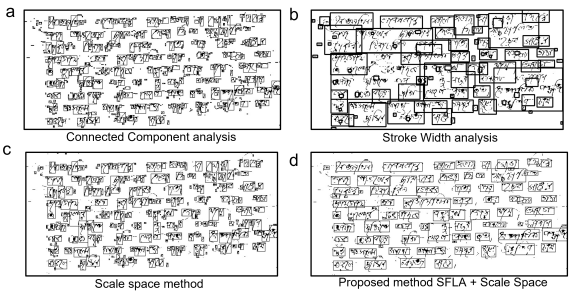

Figure 3: Segmented output images, (a - c) Segmented output images using Connected Component Analysis, Stroke Width Analysis, and Scale space method. (d) proposed SFLA-SS.

Similarly, we apply all four techniques to the BN-HTR dataset. The output of all four techniques is shown in Figure 4. Nevertheless, the modified Scale Space technique has outperformed other baseline techniques, demonstrating its robustness.

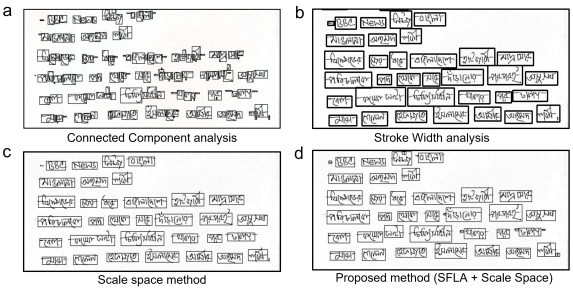

Figure 4: Compared the resulting output images after word-level segmentation on handwritten downloaded image, (a,b,c) Segmented output images using Connected Component Analysis, Stroke Width Analysis, and Scale space method. (c) Image output after segmentation using our proposed method.

During the performance analysis of the standard scale space method combined with the SFLA optimizer, we observed a particular sensitivity to skewed lines, which is a problem that is exacerbated when the spacing between lines is minimal. This often resulted in the misidentification of multiple words as a single unit. The challenge was more pronounced with our in-house dataset, which comprised handwriting from school students containing higher skewness and inconsistent spacing.

6.2 Performance of U-Net + SFLA-SS

To address the issue of skewed lines, we implemented the U-Net model for line segmentation across both datasets. The images of the segmented lines were then utilized as inputs for the subsequent stage of word segmentation. In this preliminary study, we selected ten handwritten document images and annotated the lines to generate ten corresponding mask images. After that, we trained our model on ten images and their corresponding masks. When we tested, the model achieved a 78 % accuracy. In Figure 5, we presented line segmentation output obtained from the U-Net model. Following the line segmentation, we integrated the SFLA-SS to segment words from the lines, enhancing the sophistication of our word segmentation process.

We represented an image of a quantitative analysis in Figure 6. For this analysis, we manually counted the segmented boundaries from all the output images and compared the values as per the metrics specified in Section 5.2. The in-house test dataset contains a total of 241 words in 4 images. When we applied the proposed technique to the in-house dataset, 167 words were perfectly segmented. There were 46 over-segmented, 11 under-segmented, and 18 partial-segmented words. To have a % accuracy value corresponding to a technique, we have considered only perfect-segmentation and calculated the % accuracy. So, the proposed technique achieved 69.29% accuracy on handwritten school images. Similarly, we apply the other three techniques on the same set of images and compute the values. Table 1 summarizes the result of this quantitative analysis of the in-house dataset.

In the case of the downloaded handwritten image dataset, we have a total of 236 words in 4 pages. The proposed technique performed quite well on these images. 227 words were perfectly-segmented, and 9 words were over-segmented. There were no under-segmented and partial-segmented images were presented. So, the proposed model has achieved 96.18% accuracy on downloaded handwritten datasets. Similarly, we manually count and compute these values when other techniques were applied to the dataset. Table 2 presents the values obtained from all the models we implemented in this study.

6.3 Discussion

During the performance analysis of the standard scale space method combined with SFLA optimizer, we observed a particular sensitivity to skewed lines—an issue that was exacerbated when the spacing between lines was minimal. This often resulted in the misidentification of multiple words as a single unit. The challenge was more pronounced with our in-house dataset, which comprised handwriting from school students containing higher skewness and inconsistent spacing. Nevertheless, our method outperformed other baseline techniques, demonstrating its robustness. To address the issue of skewed lines, we implemented the U-Net model for line segmentation. The images of the segmented lines were then utilized as inputs for the subsequent stage of word segmentation. In this preliminary study, we trained our model on seven images and conducted tests on two images, achieving a 78% accuracy rate on the training dataset. Although the dataset was relatively small, the model exhibited promising results when applied to the test images. Here, we would like to mention that, this was a pilot study conducted to test the effectiveness of U-Net architecture in our Bengali school-domain word segmentation task. We plan to expand this research by applying the model to a larger dataset, which is expected to refine the accuracy of the output.

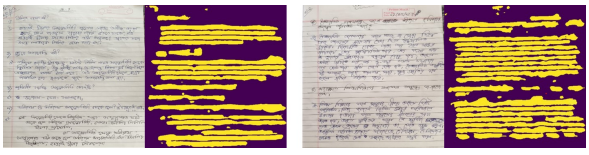

Figure 5: Images resulting from the segmentation of lines using the U-Net model.

In Figure 6, we have shown a small portion of a page to show the performance of the U-Net model. There we can see that U-Net has performed better in segmenting lines, which helped us to get a greater number of perfectly segmented words.

Segmentation Methods	Total Words	Perf. Seg.	Over Seg.	Under Seg.	Partial Seg.	Accuracy (%)
CCA	241	94	89	20	38	39
SWT	241	25	211	5	0	10.37
Scale Space	241	62	58	119	0	25.72
SFLA + Scale Space (SFLA-SS)	241	137	61	16	27	56.84
U-Net + SFLA-SS	241	167	46	11	18	69.29

Table 1: Quantitative analysis of word-level segmentation performed on handwritten in-house dataset

Segmentation Methods	Total Words	Perf. Seg.	Over Seg.	Under Seg.	Partial Seg.	Accuracy (%)
CCA	236	151	85	0	0	63.98
SWA	236	115	121	0	0	48.72
Scale Space	236	69	167	0	0	29.23
SFLA-SS	236	221	15	0	0	93.64
U-Net + SFLA-SS	236	227	9	0	0	96.18

Table 2: Quantitative analysis of word-level segmentation performed on BN-HTR dataset

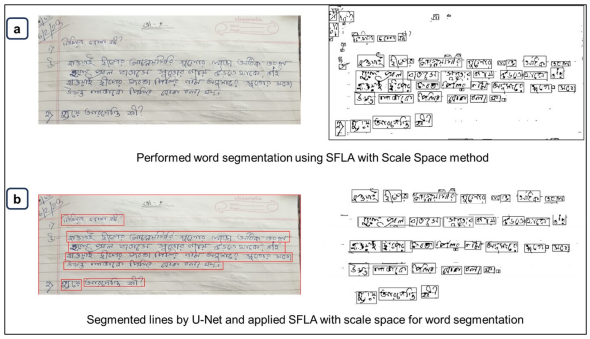

Figure 6: (a) Word level segmentation using SFLA-SS. (b) Word level segmentation using U-Net with SFLA-SS.

That text image contained 49 words. Among them, SFLA-SS segmented 18 words perfectly. On the other hand, U-Net with SFLA-SS, segmented 24 words perfectly. That clearly proves that U-Net with SFLA-SS outperforms the SFLA-SS model. We also observed that this model has shown efficacy in segmenting skewed lines, which helped to get better results.

7 Conclusion

In this study, we began by employing a non-learning-based approach for word segmentation from handwritten documents, utilizing the Shuffled Frog-Leaping Algorithm (SFLA) in conjunction with the standard Scale Space technique. This initial method proved sensitive to skewed lines and inconsistent spacing between lines. To address these issues, we subsequently incorporated a learning-based semantic segmentation approach, using the U-Net model to segment the text lines from the handwritten documents first. We then applied the SFLA-SS technique to segment the individual words from those lines. This combined technique outperformed the initial baseline method, resulting in a higher number of accurately segmented words.

Although our results have improved by adapting a traditional baseline segmentation algorithm and integrating it with the learning-based U-Net model, there is potential for further refinement. In spite of all our efforts, our final accuracy reached up to 70% only on the school dataset. However, any real application demands higher accuracy. We observed when the line is not properly maintained by the students, which is quite common in school students; for example, a phrase is written inside the gap between two lines, words are written like a superscript or subscript, a portion of the text is written with footnote, words are too close, and the overlaped words/letters the proposed method doesn't work perfectly. We need to explore the literature for a better model in such a scenario. Also, The dataset we use for training is not sufficient; we need to annotate a larger training data in order to obtain better performance of the U-Net model.

Acknowledgement

This work is supported by the Science and Engineering Research Board (SERB), India [Grant No: EEQ/2021/000687].

References

Giovanni Capobianco, Carmine Cerrone, Andrea Di Placido, Daniel Durand, Luigi Pavone, Davide Donato Russo, and Fabio Sebastiano. 2021. Image convolution: a linear programming approach for filters design. *Soft Computing*, 25(14):8941–8956.

Selman Delil, Birol Kuyumcu, and Cüneyt Aksakallı.

2021. Sefamerve ARGE at SemEval-2021 task 5: Toxic spans detection using segmentation based 1-D convolutional neural network model. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 909–912, Online. Association for Computational Linguistics.

Ziqi Dong, Hao Chang, Xuewen Pu, Peng Luo, Juhai Weng, and Zhongchen Liu. 2021. Aircraft segmentation algorithm based on unet and improved yolov4. In *2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pages 614–617. IEEE.

Muzaffar Eusuff, Kevin Lansey, and Fayzul Pasha. 2006. Shuffled frog-leaping algorithm: a memetic metaheuristic for discrete optimization. *Engineering optimization*, 38(2):129–154.

Yi Han, Ikou Kaku, Jianhu Cai, Yanlai Li, Chao Yang, Lili Deng, et al. 2013. Shuffled frog leaping algorithm for preemptive project scheduling problems with resource vacations based on patterson set. *Journal of Applied Mathematics*, 2013.

Sanasam Inunganbi, Prakash Choudhary, and Khumanthem Manglem. 2021. Meitei mayek handwritten dataset: Compilation, segmentation, and character recognition. *Visual Computer*, 37(2):291–305.

Amar Jindal and Rajib Ghosh. 2023. Word and character segmentation in ancient handwritten documents in devanagari and maithili scripts using horizontal zoning. *Expert Systems with Applications*, 225:120127.

Abhishek Khandelwal, Pritha Choudhury, Ram Sarkar, Subhadip Basu, Mita Nasipuri, and Nibaran Das. 2009. 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*, pages 369–374. Springer.

Jong-Sen Lee. 1983. Digital image smoothing and the sigma filter. *Computer vision, graphics, and image processing*, 24(2):255–269.

HR Mamatha and K Srikantamurthy. 2012. Morphological operations and projection profiles based segmentation of handwritten kannada document. *International Journal of Applied Information Systems (IJAIS)*, 4(5):13–19.

Raghavan Manmatha and Jamie L Rothfeder. 2005. A scale space approach for automatically segmenting words from historical handwritten documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1212–1225.

Raghavan Manmatha and Nitin Srimal. 1999. Scale space technique for word segmentation in handwritten documents. In *International conference on scale-space theories in computer vision*, pages 22–33. Springer.

U Pal and S Datta. 2005. Segmentation of bangla unconstrained handwritten text. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition, 2003.*

Rahul Pramanik and Soumen Bag. 2020. Segmentation-based recognition system for handwritten bangla and devanagari words using conventional classification and transfer learning. *IET Image Processing*, 14(5):959–972.

Ankur Priyadarshi and Sujan Kumar Saha. 2021. The first named entity recognizer in maithili: Resource creation and system development. *Journal of Intelligent & Fuzzy Systems*, 41(1):1083–1095.

Md Ataur Rahman, Nazifa Tabassum, Mitu Paul, Riya Pal, and Mohammad Khairul Islam. 2023. Bnhtrd: 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*, pages 1–16. CRC Press 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.

Palaiahnakote Shivakumara, Tanmay Jain, Umapada Pal, Nitish Surana, Apostolos Antonacopoulos, and Tong Lu. 2022. Text line segmentation from struck-out handwritten document images. *Expert Systems with Applications*, 210:118266.

Hongxin Xiao, Lingxi Peng, Shaohu Peng, and Yifan Zhang. 2022. Lung image segmentation based on involution unet model. In *2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pages 184–187. IEEE.