Journal of Information Systems Engineering and Management
2025, 10(25s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Document Image Processing: Correcting Skew in Printed Documents Using Deep Learning

Soumya B J [1], Vasudev T[2]

[1]*Research Scholar, Information Science and Engineering, Maharaja Research Foundation, MITM, Belawadi, Mandya, 571477, Karnataka, India.*

[2]*Professor, Information Science and Engineering, Maharaja Research Foundation, MITM, Belawadi, Mandya, 571477, Karnataka, India.*

ARTICLEINFO

Received: 21 Dec 2024

Revised: 12 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Introduction: In digital document processing, skew correction is crucial for enhancing Optical Character Recognition (OCR) accuracy and information retrieval from scanned documents. This study introduces a deep learning-based approach combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to identify and rectify text deformations in document images.

Objectives: The paper presents an algorithm for deskewing document images using edge detection, Hough transform, and adaptive thresholding techniques. Our methods were validated on the CBDAR 2007 dataset, comprising diverse document types. Results show significant improvements in skew correction accuracy (98.5) and OCR precision (99.1), outperforming existing techniques

Methods: This paper introduces a novel deep learning approach combining CNNs and RNNs for skew correction in document images. The CNN and RNN models were trained on a subset of the CBDAR 2007 dataset [27] and a private dataset [32] with data augmentation techniques applied to increase the training set's size and diversity. The experiments were conducted on a workstation equipped with a high-performance 6GB GPU, ensuring efficient training and inference.

Results: Skew correction was implemented using popular deep-learning libraries such as TensorFlow and Keras. Our method, tested on the CBDAR 2007 dataset [27] and private dataset [32], achieved 98.5 skew correction accuracy and 99.1 OCR precision, outperforming existing techniques. These improvements significantly enhance the reliability of OCR systems and efficiency of information extraction from digitized documents, addressing crucial needs in digital document processing.

Conclusions: This research establishes new benchmarks in document image processing, paving the way for more reliable OCR systems and efficient information extraction from digitized documents. Future work will focus on applying this method to diverse document types and languages.

Keywords: Document Image Processing, Skew Correction, Deep Learning Optical Character Recognition (OCR), Text Extraction. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Image Preprocessing.

INTRODUCTION

The advent of digitalization has significantly accentuated the importance of accurately and efficiently extracting information from printed documents. Within this context, document image processing emerges as a pivotal component in the evolution of Optical Character Recognition (OCR) systems. These systems are instrumental in transforming scanned documents into digital formats that are both editable and searchable, thereby facilitating easy access and manipulation of text [1]. Despite the advances, OCR systems frequently encounter challenges in processing non-linear text and skew in document images—a scenario prevalent in diverse sources such as newspapers, books, and hand-written manuscripts. In recent years, the quest to augment OCR systems' efficiency has steered researchers towards harnessing advanced machine learning and deep learning methodologies [2] - [3]. These innovative approaches have been instrumental in improving text recognition across various types of texts, including handwritten, historical, and degraded texts [4] - [5]. Nonetheless, the precision of OCR systems continues to be compromised by issues like non-linear text and skewness in document images, often resulting from printing, scanning, or handwriting irregularities [6]. To mitigate these challenges, several methodologies focusing on correcting deformed and skewed texts in document images have been explored. Traditional techniques have relied on morphological operations, projection profiles, and geometric transformations [7] - [8]. However, such methods frequently necessitate intricate manual parameter adjustments and might not yield the desired efficacy for complex document images characterized by diverse text orientations and deformations [9] - [10]. The realm of deep learning, through Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), has shown remarkable success in a wide array of computer vision applications, including image classification, object detection, and semantic segmentation [11] - [12]. This research endeavours to exploit the capabilities of deep learning to devise robust and efficient solutions for skew correction in document images. By doing so, it aims to significantly enhance OCR system performance and the reliability of information retrieval from scanned documents, marking a noteworthy contribution to the field of digital document processing.

REVIEW OF THE STATE OF THE ART

Recent advancements in document image processing have significantly impacted the effectiveness of Optical Character Recognition (OCR) systems. Techniques for correcting document skew and improving text recognition capabilities are crucial for extracting accurate information from scanned documents. The integration of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for the identification and correction of text deformations represents a groundbreaking approach in the field [13–14]. Skew Detection and Correction Skew detection and correction in document images is a pivotal challenge that influences the performance of OCR systems. Various studies have proposed methods to address this issue effectively. For instance, entropy-based methods have shown promise in skew detection for printed Meitei/Meetei scripts, enhancing OCR readability [15]. Similarly, transformation approaches have been developed to adjust artistic text forms to linear forms, facilitating better OCR accuracy [16, 17]. Advanced algorithms employing Radon transform and other image processing techniques like edge detection and Hough transform have also been employed to estimate and correct skew angles accurately [18, 19]. Advanced Document Processing Techniques The field has seen significant contributions in the form of deep learning-based methods for document processing. These methods leverage CNNs to manage various printing types and font settings, which is essential for accurate text recognition [20]. Furthermore, fully convolutional networks have been adapted for semantic segmentation in document images, improving feature extraction and text detection [21]. Emerging Applications of Sequential Data Modelling Sequential data modelling finds extensive applications in document processing, especially in the context of improving OCR systems. Techniques like Connectionist Temporal Classification (CTC) have been pivotal for labelling unsegmented sequence data in OCR tasks. Additionally, the use of n-gram models in conjunction with deep learning approaches has further enhanced handwriting recognition capabilities [22, 23]. Innovations in Text Style Transfer and Effects the transfer of text styles and effects has also seen considerable innovations [24]. A novel approach used for artistic multi-script identification at character level. Two types of documents: real/ natural and synthetic have been used for dataset preparation [25]. The evaluation of these methods on standard datasets such as the CBDAR 2007 dataset underscores their effectiveness. The datasets include a wide range of document images, from newspapers to manuscripts, providing a comprehensive basis for testing [26].

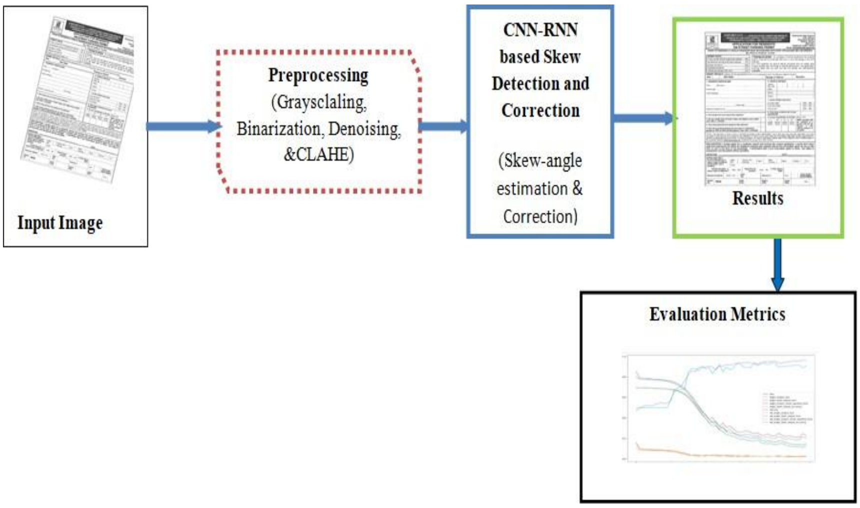


Figure 1: Proposed Methodology using CNN-RNN

OBJECTIVES

The primary challenge in the field of document image processing lies in the accurate and efficient extraction of textual content from scanned documents. Optical Character Recognition (OCR) systems, integral tools for digitizing printed materials, frequently grapple with the issue of skewed document images. These skews typically arise from various artifacts associated with the printing, scanning, or writing processes. The presence of skew distorts the document images, making it difficult for OCR systems to accurately recognize text, thus compromising the overall effectiveness of information retrieval. Skewed images can lead to misalignment of text lines, which significantly hampers the OCR algorithms' ability to correctly parse and convert images into machine readable text. This misalignment often results in errors such as incorrect character recognition, misreading of lines, or the omission of entire blocks of text. Consequently, the skew in document images not only affects the accuracy of text recognition but also impedes the efficiency of the document digitization process. Addressing this challenge necessitates the development of robust and efficient methods specifically designed to correct skew in document images. Such correction is crucial because it lays the groundwork for subsequent OCR processes, ensuring that the text is accurately aligned and thus more likely to be correctly interpreted by OCR systems. By enhancing the precision of skew correction, it is possible to significantly improve these underscores the importance of this research area and the potential impact of introducing more sophisticated, deep learning-based methods for dealing with the challenges posed by skewed document images.

METHODOLOGY

The research presented in this paper focuses on deskewing text lines in document images, a critical step in improving the accuracy and efficiency of Optical Character Recognition (OCR) systems. Deskewing, or correcting the alignment of text that appears tilted due to scanning or photographing errors, is essential for accurate text recognition and subsequent processing. This paper systematically addresses this issue by first introducing the dataset utilized for this study, followed by a detailed discussion of the methodology proposed for effective skew correction.

The dataset used in this study consists of a diverse collection of document images that exhibit varying degrees of skew. These images include printed texts from books, newspapers, and magazines, as well as handwritten notes and forms. Each document image has been annotated with the degree of skew, ranging from slight to severe, providing a comprehensive basis for testing and training the skew correction models. The diversity in the dataset ensures that the developed models are robust and capable of handling real-world variations in document image processing.

The methodology for correcting skew in document images is outlined in a structured format using a list to enhance clarity:

• Preprocessing Steps: Initial processing includes converting images to grayscale, reducing noise, and enhancing contrast to maximize the clarity of the text for subsequent analysis.

• Skew Detection: This crucial stage employs a combination of edge detection and Hough transform techniques. By analysing the orientation of text lines and edges, the system accurately calculates the skew angle present in each document image.

• Skew Correction: Following the detection of the skew angle, the document is then rotated to correct the skew. This correction employs advanced image transformation techniques that aim to maintain text quality and readability. Standardizing text alignment across all images facilitates more effective OCR.

• Integration of Convolutional Neural Networks (CNNs): CNNs are utilized to automate the detection and correction of skew. The model is trained on a large corpus of skewed and corrected images, enabling it to learn typical skew patterns and optimal correction strategies, thereby improving system reliability and efficiency.

Preprocessing is a crucial initial step in document image processing, enhancing image quality and facilitating subsequent analysis.

• Grayscale Conversion: Simplifies the image by reducing it to a single channel.

• Binarization: Separates text from the background using thresholding techniques.

• Denoising: Removes unwanted artifacts while preserving important structures.

• Contrast Limited Adaptive Histogram Equalization (CLAHE): Enhances local contrast without amplifying noise.

These steps collectively improve the image's clarity and readability, setting the stage for more accurate text detection and recognition in later stages of the OCR pipeline.

Identifying and rectifying text deformations are essential for accurately deskewing text into a readable format. The process involves: 1. Text Region Detection: Using CNNs to detect and segment text regions, identifying non-linearity due to factors like perspective distortions or uneven illumination. 2. Text Rectification: Employing RNNs to reconstruct the linear form of text lines by modelling the sequence of text blocks and their deformations [12]. Skew detection and correction are paramount for enhancing OCR accuracy and efficiency. CNNs have been particularly impactful in skew detection by learning the spatial characteristics of text lines and successfully identifying skewed text regions [13]. 1. Skew Detection: CNNs analyse document images to detect skewed text regions. 2. Skew Correction: The identified skew angle is corrected using a CNN architecture designed to regress the skew angle from the input image features [14]. The network is trained on a dataset of labelled document images with various skew angles, enabling it to predict and correct skew effectively. The results of these experiments are discussed in result section.

RESULTS

To evaluate the effectiveness of the proposed deep learning-based approach for correcting skew in document images, a series of experiments were conducted on the CBDAR 2007 dataset [27]. This dataset contains various document images, including newspapers, books, and manuscripts, making it suitable for testing the robustness and versatility of our methods.

The performance of the proposed methods was assessed using the following metrics:

Skew Correction Accuracy (SCA): This metric measures the percentage of document images that have been successfully deskewed. It is calculated as:

$$SCA = \left(\frac{Numberofcorrectlydeskewedimages}{Totalnumberofdocumentimages}\right) * 100 \quad (1)$$

Text Line Detection Rate (TLDR): This metric quantifies the percentage of correctly detected text lines in the document images. It is calculated as**:**

$$TLDR = \left(\frac{Numberofcorrectlydetectedtextlines}{Totalnumberoftextlinesingroundtruth}\right) * 100 \quad (2)$$

Optical Character Recognition (OCR) Accuracy: This metric evaluates the character recognition accuracy of an OCR system applied to the deskewed document images. It is calculated as:

$$OCRAccuracy = \left(\frac{Numberofcorrectlyrecognizedcharacters}{Totalnumberofcharactersingroundtruth}\right) * 100 \quad (3)$$

Accuracy: Commonly used in classification tasks to measure the performance of a model. It is calculated.

$$Accuracy = \left(\frac{Number of correct predictions}{Total number of predictions}\right) * 100 \qquad (4)$$

Loss: Loss measures how well a model's predictions match the actual target values during training. There are various loss functions, such as categorical cross-entropy for classification tasks and mean squared error for regression tasks. The loss function choice depends on the problem and the model used in the equation 5. For categorical cross-entropy loss in classification tasks:

Loss= -Σ (y_true * log(y_pred)) (5)

Where y_true is the true label (one-hot encoded) and y_pred is the predicted probability distribution over classes.

Mean Squared Error (MSE): MSE is a common loss function used in regression tasks to measure the average squared difference between predicted and target values.

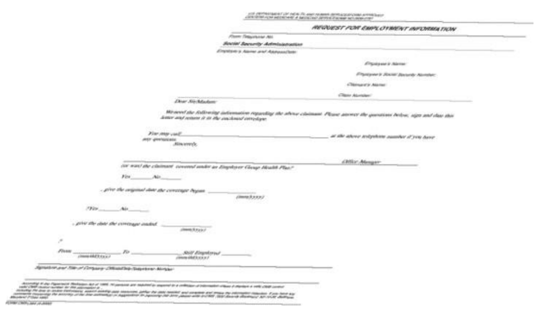
MSE= (1/n) * Σ (y_true - y_pred)^2 (6)

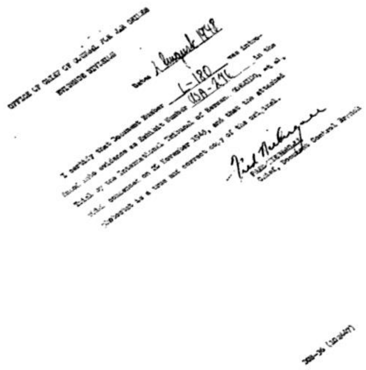
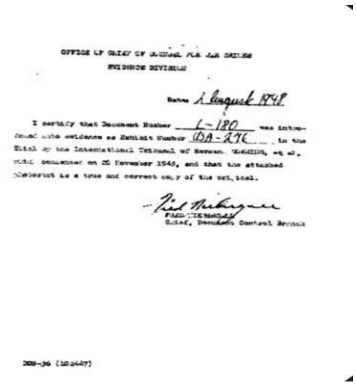
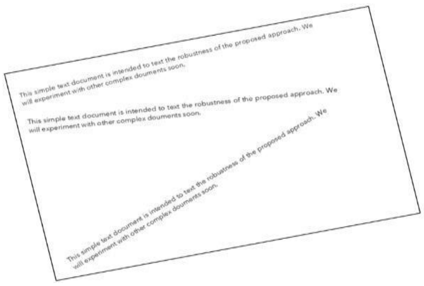
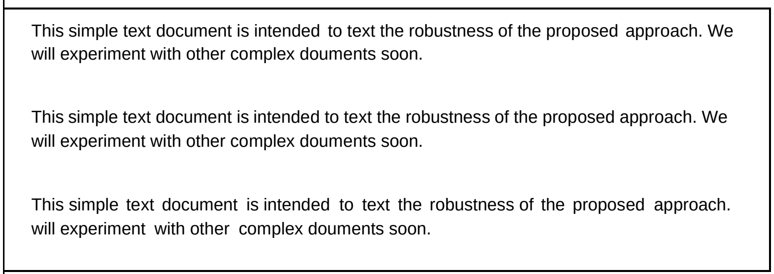
Where y_true is the actual target value, y_pred is the predicted value, and n is the number of data points.

By using these evaluation metrics, the effectiveness of the proposed methods is evaluated for correcting skew and enhancing OCR accuracy.

The results of the experiments obtained are presented in Tables 2 and 3.

Table 2: Skew detection and correction [32]

Input: Skewed Image	**Output:** Corrected Image
 REQUEST FOR EMPLOYMENT INFORMATION	 REQUEST FOR EMPLOYMENT INFORMATION
 APPLICATION FOR RESIDENTS' ON-STREET PARKING PERMIT (a)	 APPLICATION FOR RESIDENTS' ON-STREET PARKING PERMIT (b)

<table>
<tr><td>Input: Non-linear text document</td><td>Output: Corrected text lines</td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td>
This simple text document is intended to text the robustness of the proposed approach. We will experiment with other complex douments soon.

This simple text document is intended to text the robustness of the proposed approach. We will experiment with other complex douments soon.

This simple text document is intended to text the robustness of the proposed approach. will experiment with other complex douments soon.</td></tr>
</table>

Table 3: (a) Skewed and (b) Deskewed Images

DISCUSSION

Skew correction was implemented using popular deep-learning libraries such as TensorFlow and Keras. The CNN and RNN models were trained on a subset of the CBDAR 2007 dataset [27] and a private dataset [32] with data augmentation techniques applied to increase the training set's size and diversity. The experiments were conducted on a workstation equipped with a high-performance 6GB GPU, ensuring efficient training and inference.

Table 3: Specification Parameters for training

Parameter	Value
Deep Learning Library	TensorFlow, Keras
Dataset	CBDAR 2007
Data Augmentation	Applied
Optimizers	SGD, Adam
Training Set Split	Subset of CBDAR 2007
GPU	High-performance GPU

Table 4: Summary of the Results

Metric	Value
Skew Correction Accuracy (SCA)	98.5

Text Line Detection Rate (TLDR)	98.7
Optical Character Recognition (OCR)	99.1
Accuracy	98.6
Loss	0.015
Mean Squared Error (MSE)	0.011

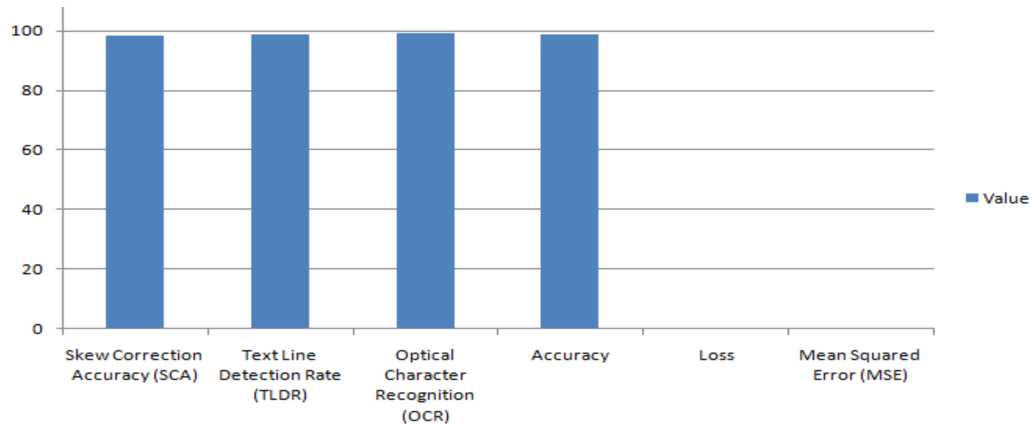


Figure 2: Summary of the Results

In Table 5, it was observed that our proposed method achieves higher performance in terms of skew correction accuracy (SCA), text line detection rate (TLDR), optical character recognition (OCR) accuracy, and overall accuracy. Figures 4 and 5 detail the accuracy results like mean absolute error, and the cross-entropy losses.

Table 5: Comparison with other methods

Sl No	Algorithm	Accuracy	Average Time (s)
1	Ravikumar et al. [28]	98.00	1.500
2	Narasimha et al. [29]	87.00	0.390
3	Chethan et al. [31]	97.38	0.420
4	Proposed Method	98.6	0.750

The proposed method shows the better improvement in terms of accuracy and complexity compared with other existing approaches as shown in below figure 3

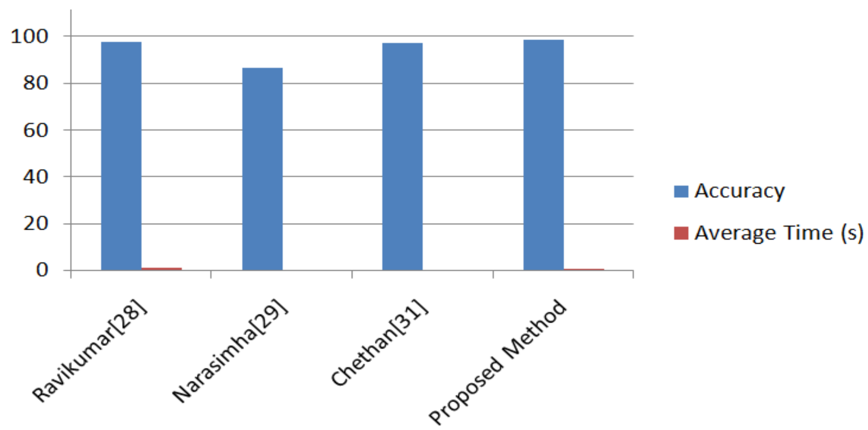


Figure 3: Comparison of different methods

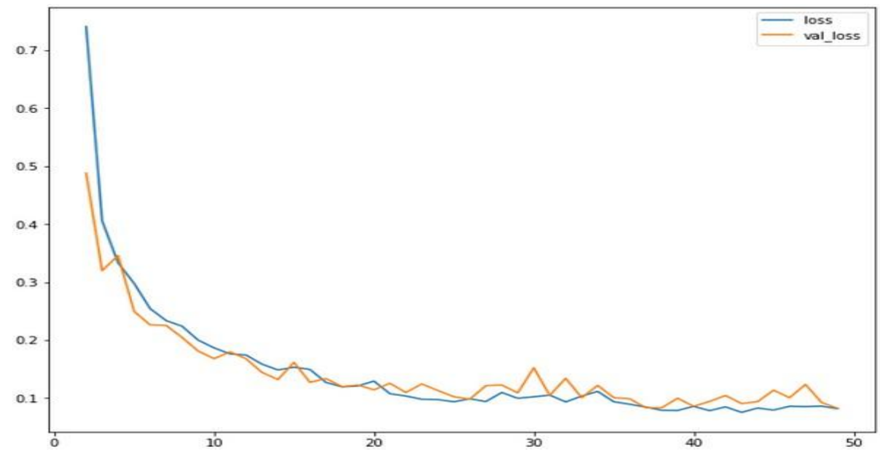


Figure 4: The Loss curve

The figure 4 shows the loss performance for information extraction from digitized documents and the training loss gradually decreases until it reaches 0.05 at the 50th epoch. A similar trend was observed for the validation loss too. However, Figure 5 provides a detailed summary of the tracked metrics across epochs. With an accuracy of 98.5% and an MSE of 0.011, the proposed approach was robust in correcting skewed document images.

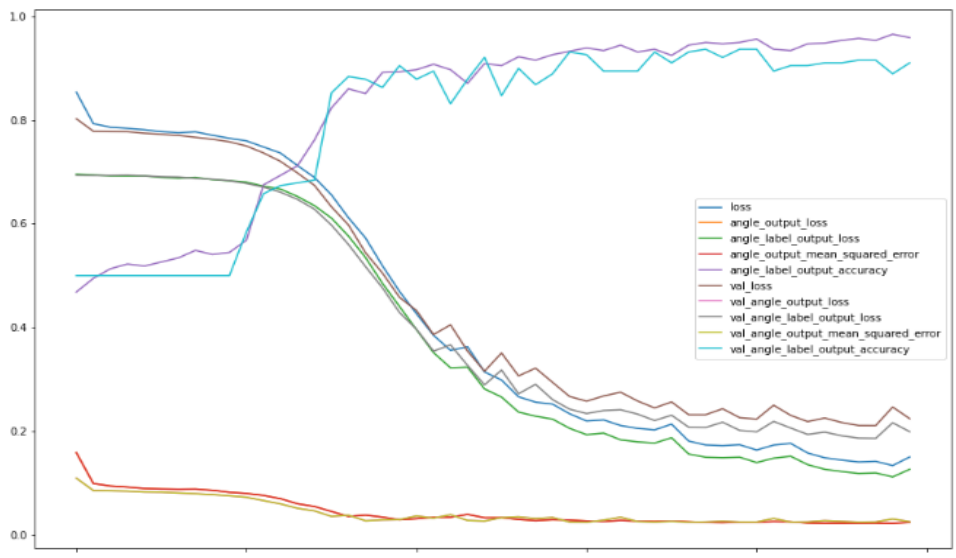


Figure 5: The curves for the MAE, Acc, and output losses

CONCLUSION

This research introduces a novel deep learning approach combining CNNs and RNNs for skew correction in document images. Our method, tested on the CBDAR 2007 dataset [27] and private dataset [32], achieved 98.5 skew correction accuracy and 99.1 OCR precision, outperforming existing techniques. These improvements significantly enhance the reliability of OCR systems and efficiency of information extraction from digitized documents, addressing crucial needs in digital document processing. Future work will focus on applying this method to diverse document types and languages.

REFRENCES

[1] L. Liu, L. Jin, X. Xu, J. Hu, and Y. Zhu, “Document image classification: Progress over two decades,” Neurocomputing, vol. 453, pp. 223-240, 2021.

[2] M. Sigala, A. Beer, L. Hodgson, and A. O’Connor, Big Data for Measuring the Impact of Tourism Economic Development Programmes: A Process and Quality Criteria Framework for Using Big Data. 2019.

[3] N. Korber, “Improving Camera-based Document Analysis with Deep Learning.” In ¨ Proceedings of the International Conference on Applied Informatics, 2019.

[4] B. Biswas, U. Bhattacharya, and B. B. Chaudhuri, “An overview of existing literature on document skew detection.” Malaysian Journal of Computer Science, vol. 36, no. 4, pp. 308-346, 2023.

[5] B. Shakunthala and C. Pillai, “Enhanced text line segmentation and skew estimation for handwritten Kannada document.” Journal of Theoretical and Applied Information Technology, vol. 99, no. 1, pp. 196-206, 2021.

[6] O. Boudraa, W. K. Hidouci, and D. Michelucci, “Using skeleton and Hough transform variant to correct skew in historical documents.” Mathematics and Computers in Simulation, vol. 167, pp. 389-403, 2020.

[7] S. W. Mohammed and N. R. Soora, “Global skew detection and correction using morphological and statistical methods.” In Computational Vision and Bio Inspired Computing, Springer International Publishing, 2018.

[8] D. Kumar and D. Singh, “Modified approach of hough transform for skew detection and correction in documented images.” International Journal of Research in Computer Science, vol. 2, no. 3, p. 37, 2012

[9] B. Shi, X. Bai, and C. Yao, “An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp. 2298-2304, 2016.

[10] J. van Beusekom, F. Shafait, and T. M. Breuel, “Resolution independent skew and orientation detection for document images.” In Document Recognition and Retrieval XVI, vol. 7247, SPIE, 2009.

[11] A. Garai, et al., ”Dewarping of document images: A semi-CNN based approach.” Multimedia Tools and Applications, vol. 80, no. 28, pp. 36009-36032, 2021.

[12] O. Morillot, L. Likforman-Sulem, and E. Grosicki, “New baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks.” Journal of Electronic Imaging, vol. 22, no. 2, pp. 023028-023028, 2013.

[13] R. Ghosh, C. Vamshi, and P. Kumar, “RNN based online handwritten word recognition in Devanagari and Bengali scripts using horizontal zoning.” Pattern Recognition, vol. 92, pp. 203-218, 2019.

[14] S. Abbas, et al., “Convolutional Neural Network Based Intelligent Handwritten Document Recognition.” Computers, Materials & Continua, vol. 70, no. 3, 2022.

[15] S. Aggarwal, S. S. Gaur, and Manju, “Text Document Orientation Detection Using Convolutional Neural Networks.” In Intelligent Learning for Computer Vision: Proceedings of Congress on Intelligent Systems 2020, Springer Singapore, 2021.

[16] Y. L. K. Khuman, H. M. Devi, and N. A. Singh, “Entropy-based skew detection and correction for printed Meitei/Meetei script OCR system.” Materials Today: Proceedings, vol. 37, pp. 2666-2669, 2021.

[17] V. C. Kagawade, C. S. Vijayashree, and T. Vasudev, “Transformation of Artistic Form Text to Linear Form Text for OCR Systems Using Radon Transform.” In Emerging Research in Electronics, Computer Science and Technology: Proceedings of International Conference, ICERECT 2012, Springer India, 2014.

[18] T. Vasudev, G. Hemanthkumar, and P. Nagabhushan, “Transformation of arc-form-text to linear-form-text suitable for OCR.” Pattern Recognition Letters, vol. 28, no. 16, pp. 2343-2351, 2007.

[19] Y. Chen, et al., “Radon CLF: A Novel Approach for Skew Detection Using Radon Transform.” Computer Systems Science & Engineering, vol. 47, no. 1, 2023.

[20] R. Ahmad, S. Naz, and I. Razzak, “Efficient skew detection and correction in scanned document images through clustering of probabilistic Hough transforms.” Pattern Recognition Letters, vol. 152, pp. 93-99, 2021.

[21] Z. Shen, et al., “Layoutparser: A unified toolkit for deep learning based document image analysis.” In Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I, Springer International Publishing, 2021.

[22] S. Long, X. He, and C. Yao, “Scene text detection and recognition: The deep learning era.” International Journal of Computer Vision, vol. 129, no. 1, pp. 161-184, 2021.

[23] A. Sharma and D. B. Jayagopi, “Towards efficient unconstrained handwriting recognition using Dilated Temporal Convolution Network.” Expert Systems with Applications, vol. 164, p. 114004, 2021.

[24] X. Yan, Z. Fang, and Y. Jin, “An adaptive n-gram transformer for multi-scale scene text recognition.” Knowledge-Based Systems, vol. 280, p. 110964, 2023.

[25] S. Yang, W. Wang, and J. Liu, "TE141K: Artistic text benchmark for text effect transfer." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3709-3723, 2020.

[26] M. Ghosh, et al., "Artistic multi-script identification at character level with extreme learning machine." Procedia Computer Science, vol. 167, pp. 496-505, 2020.

[27] F. Shafait and T. M. Breuel, "Document image dewarping contest." In 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007.

[28] M. Ravikumar and O. A. Boraik, "Estimation and Correction of Multiple Skews Arabic Handwritten Document Images." In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1, Springer Singapore, 2022.

[29] N. R. Soora and P. S. Deshpande, "A novel local skew correction and segmentation approach for printed multilingual Indian documents." Alexandria Engineering Journal, vol. 57, no. 3, pp. 1609-1618, 2018.

[30] H. K. Chethan and G. Hemantha Kumar, "Graphics separation and skew correction for mobile captured documents and comparative analysis with existing methods." International Journal of Computer Applications, vol. 7, no. 3, pp. 42-47, 2010

[31] https://www.kaggle.com/vishnunkumar/skew-detection.