Grenze International Journal of Engineering and Technology, June Issue

Optimizing OCR Accuracy for Folded and Mirrored Text Documents

Vedant Chavan[1], Vaishnavi Kadukar[2], Isha Patade[3] and Ramchandra Mangrulkar[4]
[1-4]Dept. of Information Technology, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
Email: vedantchavan29@gmail.com, vgkadukar21@gmail.com, ishaapatade14@gmail.com,
ramchandra.mangrulkar@djsce.ac.in

***Abstract*— This literature review explores methods to enhance OCR performance on inverted, reversed, or backend text, addressing the limitations of conventional OCR systems in accurately recognizing such orientations. The methods used range from simple image processing techniques like edge detection and histogram equalization to advanced approaches such as convolutional neural networks, support vector machines, fuzzy logic, and neural networks, addressing uncertainties in character shapes, angular orientations, and feature alignments. This paper discusses the use of the TrOCR model for recognizing text on the backside of documents, addressing issues such as ink bleed-through, alongside preprocessing and postprocessing solutions. It evaluates accuracy, highlighting the model's utility in enhancing OCR dependability. The review underscores the need for more robust algorithms capable of handling diverse document types and text orientations while preserving the originality of documents for digitization and archival purposes.**

***Index Terms*—Optical Character Recognition (OCR), Backside Text Recognition, Document Image Processing, TrOCR (Transformer-based OCR), Image Enhancement Techniques, Image Thresholding, Text Extraction from Image.**

I. INTRODUCTION

Historical or physically degraded texts are difficult to extract with accuracy, especially when dealing with cultural, legal, and historical records. Physical degradation of documents like legal contracts, manuscripts, or even ancient records often occurs after years of aging, exposure, and handling. This degradation may take different forms, such as ink fading, tears, and stains and more importantly, folds and creases, causing deformations which distort the alignment and readability of the text. Most conventional Optical Character Recognition (OCR) systems work best with well-preserved and flat text [1] and are, therefore, incapable of addressing these distortions.

This study aims to design and develop an OCR-based text extraction system for reading documents with double folds or severe degradation. The proposed system will use sophisticated algorithms in enhancing the image along with OCR techniques in order to improve the accuracy and recovery of the text. It will apply state-of-the-art preprocessing techniques like edge detection, histogram equalization, and dynamic thresholding to enhance the quality of degraded images, thus making texts more readable and visible.

Often, the libraries, museums, and archives are involved with rare, fragile documents that cannot be handled frequently.

Grenze ID: 01.GIJET.11.2.37_20

Sometimes for their historical or legal value, these artifacts are at the verge of deteriorating or even loss if not properly conserved. In such instances, an OCR that can read and understand text on broken, folded, or partially damaged documents would be highly useful. It would ensure that information of value is not missed and that preservation in digital formats does not damage the physical integrity of a document.
Several existing works have focused on similar objectives [2] by devising algorithms for detecting mirrored text in natural images, mitigating the effect of flipped labels in scanned medical slides, or improving the performance of OCR systems within everyday use. These techniques usually involve RGB-to-grayscale conversion, noise removal, and uses machine learning for classification and identification.
This review will critically examine these existing methods to identify their respective strengths and limitations based on their ability to manage text in damaged or folded documents. The present gap in research will be highlighted and a framework proposed for building a more efficient OCR system for adverse document conditions.

II. Related Works

To deal with orientation in images, [3] introduces a system for classifying slide labels into three categories: it has 3 basic orientations, which are normal, a flipped or upside down and a blank orientation. In its design, it takes into consideration two major factors; the time taken for the computation process and the accuracy in the detection of orientation. Due to the correct orientation of the slides, the system increases the effectiveness of scanning operations and reduces the possibility of errors and further analysis of the data. The method combines the blurring image, thresholding, and denoising with OCR to produce a word list with confidences. Classification decisions in the system are made based on the number of text boxes and the certainty of the recognized words. In [4], the use of OCR is employed to extract key data from printed bills and invoices containing total values, date and itinerary information. Several pre-processing steps are then performed on the image such as denoising and edge detection so as to extract the bill from the image. After pre-processing, the image is processed by Tesseract OCR to extract the text, using algorithms for text segmentation for different styles, and fonts and multiple languages. An Android application was developed for uploading bills in form of images and enhance the extraction accuracy as the manually-entered values were compared with the extracted values.
Kwak et al. [5] presents an approach for distortion correction of misoriented text in document images, which is an important preprocessing step carried out in OCR tasks. The approach deals with focal directional gradient features computed in four directions (0°, 90°, 180°, 270°) on the grayscale images. These gradients are analysed using features such as the sum of the values in the array and kurtosis which leads to the classification of text orientation in multi-level decision tree.
The researchers tried this method using English and Telugu text and found it at 94% accuracy but it was troublesome in handling low-resolution images, graphics-rich documents, and multiple fonts. However, it provides a tentative solution for correcting textual orientation with no requirements of extra hardware and can be extended to apply to a wider range of vocabulary and structures in future OCR developments.
Another method proposed in [6] is a newly developed binarization technique for enhancing the quality of the distorted document images. The technique is mainly intended for a general treatment of probable image degradation like ink bleeding, uneven lighting, smudges, blot, shadows and effectively managing a complex background.
The proposed algorithm involves four steps: generating a Laplace-domain pixel value image, creating a page-pointer map, regional selection, and fine-tuning through post-processing. It proves that the proposed method has higher accuracy in comparison with several adaptive binarization techniques. But improvements are still required to increase its performance for such kinds of images especially those with swift contrast.
In [7], an innovative algorithm to perform OCR on both sides of a page without flipping the page together with a method that can handle shadow text when the actual text is difficult to see is presented. This technique is more effective for palm leaves, manuscripts, and other records which cannot be scanned by normal OCR techniques. The method uses Neural Fuzzy Hybrid System of Artificial Neural Networks and Fuzzy Logic Controllers which improves the accuracy of recognition.
The system employs optimized resolution, contrast, and brightness to capture text that contains both front and partial back text. Neural networks recognize characters and their shapes while fuzzy logic considers variations in the shades of ink, velocity, and orientation. This is regarding the removal of overlapping or blended text in the document to make layers conspicuous, and enhancement in text clearness through managing bounder difficulties. Thus, this hybrid approach is well-suited for complex documents with textual variations and orientations, as well as degraded documents.

Text-DIAE is a self-supervised Degradation Invariant Autoencoder for Text Recognition and Document Enhancement [8]. It enhances OCR, handwritten text recognition, and document enhancement by learning visual representations without the need for labelled data with the help of a transformer neural network architecture. Masking, blurring, and noise addition preprocessing tasks represent document degradation accurately and help the model learn meaningful visual concepts.

By comparing with previous approaches, the proposed method, Text-DIAE, achieves better results on handwritten text recognition and OCR in deblurred documents. The authors posit that pretext tasks must match downstream tasks, and they call for more exploration into the phenomenon.

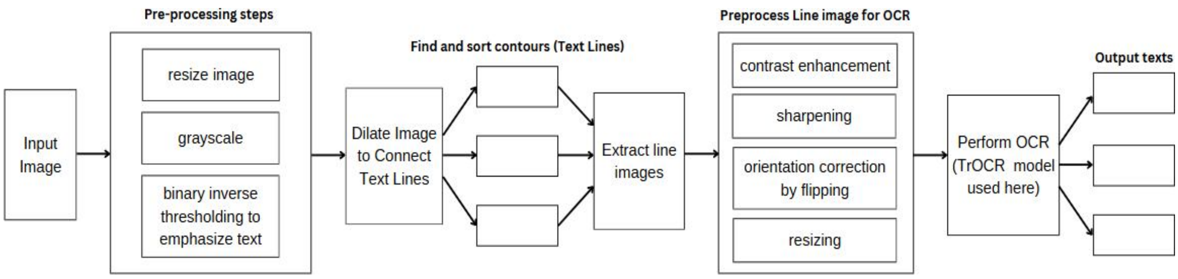

Figure 1. Flowchart representing the step-by-step process of the proposed system

III. Proposed Methodology

The method developed in this research is designed to identify the locations of handwritten text in old documents, as well as separate them using Image Processing and Optical Character Recognition (OCR). It consists of several phases, such as document preprocessing [9], line segmentation, eliminating various difficulties that arise from varying text formats, document quality and mirrored text, and text recognition phases, resulting in a system that achieved 90% accuracy.

A. Image Preprocessing

The image of the document is initially in RGB color format, which is more detailed to the colors important for preprocessing. Due to the high resolution of the scanned images, the width of images larger than 1000 pixels is reduced, and their height is proportionally adjusted.

To convert the document into a binary form, the first pre-processing method performed is grayscale conversion, which is then followed by binary thresholding as revealed in the literature [10]. With the chosen threshold value, which is 80, the image is inverted such that the text regions appear white on a black background. This contrast enhancement separates the text so that the subsequent line segregation becomes more accurate.

B. Line Segmentation

Algorithm 1: Process Image and Save Lines

1. **Input:** I_{path} (Path to input image), O_{dir} (Output directory)
2. **Output:** L
3. Load image and convert to RGB: $I_{RGB} \leftarrow I_{BGR}$
4. $(h, w, c) \leftarrow S(I_{RGB})$
5. **if** $w > 1000$ **then**
   $I_{RGB} \leftarrow \text{Resize}(I_{RGB}, (1000, 1000\ w/h))$
   end if
6. $I_{thresh} \leftarrow \text{Threshold}(I_{RGB})$
7. Kernel $k \leftarrow \mathbb{1}_{3\times85}$: $I_{dilated} \leftarrow D(I_{thresh}, k)$
8. Find contours and sort by y-coordinate:
   $C_{sorted} \leftarrow \text{sortedContours}(C, y_{top})$
9. Initialize an empty list of lines: $L \leftarrow []$
10. **for** each $C_i \in$ Csorted **do**
    $I^i_{line} \leftarrow I_{RGB}[y_i : y_i + h_i, x_i : x_i + w_i]$
    L.append(I^i_{line})
    end for
11. **return** L

Algorithm 2: Preprocess Line Image for OCR

1. **Input:** I_{line_path} (Path to line image)
2. **Output:** $I_{processed}$ (Pre-processed image for OCR)
3. Load the grayscale image
4. Apply CLAHE: $I_{CLAHE} \leftarrow C(I_{line})$
5. $I_{blurred} \leftarrow G(I_{CLAHE}, (9, 9), 10.0)$
6. $I_{sharpened} \leftarrow 1.5 \times I_{CLAHE} - 0.5 \times I_{blurred}$
7. $I_{binary} \leftarrow T_{Otsu}(I_{sharpened})$
8. $I_{dilated} \leftarrow D(I_{binary}, \mathbb{1}_{2\times2})$
9. Flip image: $I_{flipped} \leftarrow$ cv2.flip($I_{dilated}$, 1)
10. Resize image: $I_{resized} \leftarrow R(I_{flipped}, 0.8)$
11. Convert to RGB: $I_{processed} \leftarrow I_{resized}$
12. **return** $I_{processed}$

Dilation is then performed on the binary image using a horizontal kernel to join the character tags into proper segments of text lines. This operation helps to select the entire contour line rather than partial characters, as is appropriate for line-level OCR [11].

By applying contour detection from OpenCV and the connected regions that resulted from dilation, each line segment is detected. The contours are then sorted on the basis of the Y-coordinate within the document, which maintains readability of the text, and the recognition moves in natural reading order. Every contour bounding box is utilized as a mask over the individual line images extracted over the original document for the purpose of text extraction.

C. Enhanced Text Processing for OCR

Once each line is isolated [12], further preprocessing steps are used to optimize the images specifically for OCR accuracy:

CLAHE (Contrast Limited Adaptive Histogram Equalization): CLAHE is applied to each line image to improve the distinction of the text and reduce blurring factors by amplifying the contrast in the low-contrast zones that are often typical when working with worn-out or aged documents [13].

This local histogram equalization improves legibility, especially in the areas of the scanned document where ink may have worn off erratically.

Unsharp Masking for Detail Enhancement: To amplify fine text features, a technique known as unsharp masking [14] is used. CLAHE is performed to enhance the overall contrast of the image, and a Gaussian blur is applied to the CLAHE-enhanced image to remove noise and thin down the character edges further to better serve for OCR. It makes the text much more legible, especially when you are dealing with handwriting that is ambiguous or barely visible.

Otsu's Thresholding for Binarization: Following the sharpening step, a global thresholding technique known as Otsu's method [15] is used to threshold the image in order to segment it into foreground and background for more effective binarization of the form across the varying brightness levels of the document. This adaptive binarization helps to keep the contrast between text and background constant for each line so that text features can be seen by the OCR model.

Dilation and Mirroring for Flipped Text: The minor dilation [16] is also made to close any minor gaps in the set of characters. For recognizing mirrored text, each line is horizontally mirrored, which caters to the fact that in some historical text, due to scanning or layout, the text may appear completely reversed.

Resizing for OCR Compatibility: The line image, which has been processed in the previous steps, is then rescaled to the dimensions that are preferable for the input of the OCR model and increases OCR accuracy due to no blurring at the preferred resolution.

D. Text Recognition using TrOCR Model

Each line is processed with the TrOCR processor and VisionEncoderDecoderModel, pre-trained for the hand-written text recognition task. The image, which is preprocessed, has to be converted to the compatible pixel values in order to feed the OCR model and extract the text from it.

[17] This is achieved by decoding the pixel values of each line segment into a recognized character for compilation into readable text by the model. The output is then decoded and cleansed to provide a neatly written transcription for each line said.

This training makes the TrOCR model [18] accurate in handwriting text recognition, especially for historical documents, as most of them are ambiguous in writing style and ink quality. The recognized text of every line is collected and arranged one by one to retain the structure of the document. This compiled text includes the translation of the document's content into machine-readable form, where the content and the layout order were preserved.

Algorithm 3: OCR Detection
• Input: Iprocessed (Pre-processed image for OCR)
• Output: Tgenerated (Detected text)
• Load the TROCR processor and model
• Ipixels ← P (images = Iprocessed)
• Generate text: Tgenerated ← generate(Ipixels)
• Tdecoded ← batch decode(Tgenerated)
• return Tdecoded

IV. RESULTS

From this project, it is evidenced that it can accomplish the extraction of text on folded or degraded documents from images. The text which was invisible or had very low contrast becomes much cleaner for the system and therefore easier to read when processed with a mixture of CLAHE, unsharp masking, Otsu thresholding functions from the OpenCV library. In the preprocessing stage, the text is separated by contour detection, then prepared for recognition through the TrOCR model that easily recognizes distorted handwritten text. The output in terms of text produces a clear and accurate interpretation of the content of the document, which is more useful in the case of indexes and text data.

These operations rightly handle poor document conditions such as poor contrast or mirror image writing so as to harvest good text results from such trials. The extracted lines of the text appear to be in order thus producing a neat and clean output. These results ensure that this system is quite robust to deliver useful applications such as archiving research in damaged or historical documents, document restoration, and digital data extraction out of damaged or historically preserved documents.

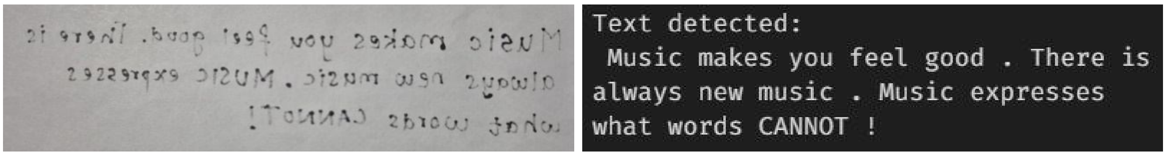

Figure 2. Backside image of a paragraph used for OCR evaluation and the model output for the image

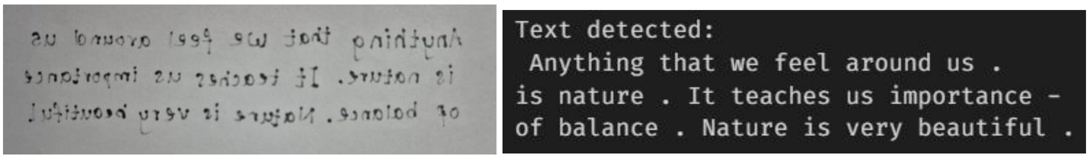

Figure 3. Backside image of a paragraph consisting of long words and the model output for the image

TABLE I. ACCURACY RESULTS FOR DIFFERENT CATEGORIES

Category	Correctly Identified	Accuracy (%)
Letters	21/26	80.77%
Words	12/18	66.67%
Numbers	5/10	50.00%
One-line	8/10	80.00%
Paragraphs	3/4	75.00%

V. CONCLUSIONS

The authors can conclude on the topic discussed and proposed. Future enhancement can also be briefed here.

This research shows how an OCR based text extraction system that is capable of effectively and efficiently extracting text from degraded and fold document can be designed and implemented. Applying techniques like CLAHE, unsharp masking, and Otsu thresholding, the system increases text contrast and makes text areas visible even if it is low contrast and mirrored text. The ability to utilize a pre-existing deep learning model, TrOCR, initially developed for recognition of handwritten text increases the effectiveness of the system to properly transcribe more complex historical documents as such documents reveal notable instances of questionable quality writing and inconsistent ink density. This approach reflects the integrity of the typical document structure as well as helps in document digitization and storage, making the extracted text available for further analysis. Further work could be dedicated to fine-tuning described system for diverse languages as well as expanding the scope of the presented system for degradation types and applying the methodology for other document conservation and digitization projects.

REFERENCES

[1] Parul Sahare and Sanjay B. Dhok. Review of text extraction algorithms for scene-text and document images. IETE Technical Review, 34(2):144–164, 2017.

[2] F. Drira. Towards restoring historic documents degraded over time. In Second International Conference on Document Image Analysis for Libraries (DIAL'06), pages 8 pp.–357, 2006.

[3] Aditya Saxena and Ruchi Verma. Analysis of flip-mirror slide labels-label classification using image morphing techniques. (1), 2023.
[4] Alan Jiju, Shaun Tuscano, and Chetana Badgujar. Ocr text extraction. International Journal of Engineering and Management Research, 11(2):83–86, 2021.
[5] Shobha Rani, Vijaya Shree Cs, and T Vasudev. An unsupervised classification technique for detection of flipped orientations in document images. International Journal of Electrical and Computer Engineering, 6(5):2140, 2016.
[6] Usha Rani, Amandeep Kaur, and Gurpreet Josan. A new binarization method for degraded document images. International Journal of Information Technology, 15(2):1035–1053, 2023.
[7] Santosh Kumar Henge and B Rama. Ocr-mirror image reflection approach: Document back side character recognition by using neural fuzzy hybrid system. In 2017 IEEE 7th International Advance Computing Conference (IACC), pages 738–743. IEEE, 2017.
[8] Mohamed Ali Souibgui, Sanket Biswas, Andres Mafla, Ali Furkan Biten, Alicia Forne´s, Yousri Kessentini, Josep Llado´s, Lluis Gomez, and Dimosthenis Karatzas. Text-diae: a self-supervised degradation invariant autoencoder for text recognition and document enhancement. In proceedings of the AAAI conference on artificial intelligence, volume 37, pages 2330–2338, 2023.
[9] Alan Koshy, Niranj Balakumar M.J., Shyna A., and Ansamma John. Preprocessing techniques for high quality text extraction from text images. In 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), pages 1–4, 2019.
[10] Panca Mudjirahardjo. The effect of grayscale, clahe image and filter images in convolution process. International Journal of Advanced Multidisciplinary Research and Studies (IJAMRS), 4(3):943–946, 2024.
[11] Vassilis Papavassiliou, Vassilis Katsouros, and George Carayannis. A morphological approach for text-line segmentation in handwritten documents. In 2010 12th International Conference on Frontiers in Handwriting Recognition, pages 19–24. IEEE, 2010.
[12] Veronica Romero, Joan Andreu Sanchez, Vicente Bosch, Katrien Depuydt, and Jesse de Does. Influence of text line segmentation in handwritten text recognition. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pages 536–540. IEEE, 2015.
[13] Hari Kumar Buddha, James Stephen Meka, and Praveen Choppala. Ocr image enhancement & implementation by using clahe algorithm. Mukt Shabd J, 9:3595–3599, 2020.
[14] A. Polesel, G. Ramponi, and V.J. Mathews. Image enhancement via adaptive unsharp masking. IEEE Transactions on Image Processing, 9(3):505–510, 2000.
[15] Oliver Nina, Bryan Morse, and William Barrett. A recursive otsu thresholding method for scanned document binarization. In 2011 IEEE Workshop on Applications of Computer Vision (WACV), pages 307–314, 2011.
[16] Guillaume Renton, Yann Soullard, Cle´ment Chatelain, Se´bastien Adam, Christopher Kermorvant, and Thierry Paquet. Fully convolutional network with dilated convolutions for handwritten text line segmentation. International Journal on Document Analysis and Recognition (IJDAR), 21:177–186, 2018.
[17] Hongkuan Zhang, Edward Whittaker, and Ikuo Kitagishi. Extending trocr for text localization-free ocr of full-page scanned receipt images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, pages 1479–1485, October 2023.
[18] Filipe Lauar and Valentin Laurent. Spanish trocr: Leveraging transfer learning for language adaptation. arXivU preprint arXiv:2407.06950, 2024.