

SEGMENTATION OF NON-TEXT FROM BILINGUAL REAL-TIME OFFICE DOCUMENT IMAGES USING U-NET ARCHITECTURE

SHIVAKUMAR G¹, RAVIKUMAR M², SAMPATHKUMAR S³ and

SHIVAPRASAD B J⁴

^{1,2,3}Department of Computer Science, Kuvempu University, Jnanasahyadri, Shivamogga, India.

Email: ¹g.shivakumarclk@gmail.com, ²ravi2142@yahoo.co.in, ³sampath1447@gmail.com

⁴Department of Computer Science and Engineering, Srinivasa Institute of Technology, Mangalore, India.

Email: shivaprasad1607@gmail.com

Abstract

In this work, we have presented an efficient approach for segmentation of non-text document information from real time office document images which are bilingual using a machine learning approach i.e., U-net architecture for experimentation purpose. We have created our own dataset containing 200 document images. Initially pre-processing is applied on the input document images proposed method is compared with other existing methods and obtained accuracy of 99% different performance measure i.e., (Specificity, Sensitivity, Precision, F1-Score) used in the experimentation.

Keywords: Document Images; Pre-Processing; Filtering; Segmentation (U-Net).

1. INTRODUCTION

The concept of document image processing, being the separation of text and non-text from scanned printed bilingual document images is a critical component of facts processing and provides inspiration for image analysis. The reliability of separation in an image-based completely digital identification system is carefully examined, with the exception of the statistics of the images. For maximal real-time record image processing, non-text portions of images must be examined promptly and properly in order to improve pooling and timing accuracy, while at the same time reducing rejection rates and increasing image first rates, by using spatial, frequency, and fuzzy filters to remove undesirable items. Document image processing has recently received the attention of researchers and strategies due to its good sized ability in possible program. A number of data extensions are used to increase the generalization capacity of the network while training it on this dataset. In addition, create a comprehensive own dataset that covers a variety of real-world conditions. A quantitative and qualitative evaluation of the proposed model must be conducted in conjunction with the previous non-learning of the basic method.

The complexity of document images is increasing, and the requested arrangements are changing among various languages, styles, font sizes, and shades, all of which make a significant difference in specific applications [1]. The scanning procedure of the text images mainly used a combination of probabilistic models, Q-tests, and PP methods [2]. Ensure that computational complexity and accuracy are balanced. One of the most crucial jobs in report picture evaluation is segmenting the numerous additives of a report image, such as text, printed

document images, and half-tones. This lets in OCR photos to be vectorized and half-tones to be compressed. Noise elimination, skew detection and correction, segmentation, feature extraction, and category are all tiers of a universal person reputation system [3]. Optical Character Recognition (OCR) devices can struggle to guarantee that the scanned report will not be skewed at some point during the optical scanning process. As an end result, de-skewing the document images is essential for enhancing the accuracy of the textual content recognition task. Many file processing algorithms are primarily based on the presumption that the report is skewed [4] [5][6].

To detect significantly changed perspectives, the immediate approach worked similarly to the Hough transform. Because normal distribution enhances reliability, authors compared the Hough transform, cross-correlation, K-nearest neighbour transform, and Fast Fourier transform. A Fast Fourier Transform improves performance and accuracy over other techniques by correcting skew in documents [7]. Handwritten documents with multiple orientations require additional pre-processing for segmentation subsequent phases for proper functioning in the handwriting recognition system [8].

A deep learning approach to detection Distorted scans document angles with different spellings language [9]. RLSA algorithm is used Rows and columns of document images [10]. Web page extraction, baseline extraction, format evaluation, or a couple of illustration and image extraction typologies. The author suggests using U-net document images and CNN-based pixel-sensible challenge-based post-processing blocks [11]. Documents are conventionally scanned using an expensive, non-portable flatbed scanner device. With the increasing popularity of mobile cameras, taking images physical documents has become the simplest way to digitise physical documents. Images are further filtered after capture by text detection and identity pipelines for content analysis and information extraction [12].

Using FFT (Fast Fourier transform) median filtering, the author proposed skew detection and correction method [13]. Skew detection and correction for Mushaf Al-Quran image pages based totally on Hough rework technique [14]. Traditional segmentation approaches, machine learning segmentation, as well as computational intelligence segmentation are indeed the three types of background subtraction. Traditional segmentation techniques include area-based segmentation, edge-based segmentation, and threshold-based approaches. Machine learning-based segmentation approaches, which are a subset of machine learning, comprise neural networks with layers for segmentation after unsupervised or supervised learning methods [15-16].

In handwritten Kannada documents with no constraints, a desk wing algorithm leads to line and word segmentation [17]. In order to determine whether skew detection is possible, the author evaluates three commonly applied techniques, namely (i) Projection Profile Analysis (PP), (ii) Hough Transform (HT), and (iii) Nearest Neighbour (NN) [18-19]. Using a geometrical method, this paper shapes a line from the components separated for various reasons [20]. A convolutional neural network (CNN) is applied to create the U-Net architecture. An algorithm for segmenting text content lines based on deep learning [22-23]. Using an Adaptive U-Net Architecture for Text Line Segmentation [24].

The boundary growing method using the linear regression evaluation (LRA) to extract the lowermost and uppermost coordinates of pixels in order to evaluate the skew perspective of a skewed document [25]. In the U-Net, huge accessible fields are enabled via down sampling operators, and we empirically select max-pooling operators rather than stride-convolutions. For the up sampling of the concatenated alerts, author has used 4×4 transposed-convolution layers with stride 2 [26]. The line segments of each institution are processed to discover a longest linked line that is decided to be the Shirrekha. The method is using while compared to Hough transform primarily based line detection method further to being strong to noise and The U-Net document image segmentation is a deep learning network architecture which is usually used for semantic segmentation tasks [27-28].

The process to separate text and non-textual areas in such an image by combining Wavelet-based Gray Level Co-Occurrence Matrix (GLCM) functions and K-method clustering [29]. In this paper, the author uses Symlet wavelet and 2-suggest classification for text segmentation from image documents [30]. This paper analysed the classification and segmentation of non-text blocks in documents into tables, graphs, and figures. Algorithms end up extra green, strong, and concise. There are numerous true segmentation and distorts estimating and correcting algorithms inside the literature overview. However, the time required to calculate the skew attitude remains a problem [31-34]. Deep Neural Networks (DNN) [35]. This research utilizes the concept of semantic segmentation with the aid of a multi-scale convolutional neural network community [36].

In ANN research, there is still a wave originating from Convolutional Neural Networks (CNN), which is broadly recognized [37]. We have Achieved impressive performance with various image segmentations Tasks and real time printed bilingual document grouped into five categories, including: as: CNN and FCN, RNN, R-CNN, Extended CNN, Attention Base Includes models, generative models and hostile models other [38].

Enhancement of real time office document images helps in better segmentation that also helps in detecting text and non-text of document images. We can use spatial or frequency domain enhancement techniques to enhance an image. It helps in removing noise, improving contrast, and preserving edges. After enhancement, the next important step is segmentation, where different conventional segmentation approaches such as area-based segmentation, edge-based segmentation, and threshold-based segmentation approaches, machine learning algorithms using supervised and unsupervised learning methods, and deep learning algorithms such as convolutional neural networks (CNN) and deep learning algorithms are used CNN (DCNN). U-Net is one such deep convolutional neural network that was developed primarily to segment office document images [39-46].

For better understanding, the remaining document part is organized as follows: section 2 discusses the related work, proposed methodology is detailed in section 3, followed by result and discussions in section 4, and finally conclusion is given in section 5.

2. RELATED WORK

Document images can be segmented accurately using several techniques. The segmentation of document images can be accomplished using a variety of techniques. An analysis of gradients and fast skew angles [2]. By using GLCM and K-means clustering, segment the document image into text and non-text regions. U-Net and ResU-Net fashions are used to separate text from Devanagari file images using Text Separation from Document Images, comparing a variety of conventional and deep-mastering techniques. There are three methods for rotating binary image documents, PP, HT, and Nearest Neighbour (NN), each to rotate the document at a different angle (document skew). An overview of four deep learning and machine learning architectures for character recognition, namely Support Vector Machines, Artificial Neural Networks, Nave Bayes, and Convolutional Neural Networks [8] [9] and Convolutional Neural Networks [10]. There are several types of Deep Learning Building Blocks. These include segmentation evaluation and extraction, baseline extraction, layout analysis, and post-processing that inhibits pixel-wise predictors based on CNN [11].

A convolutional neural network (FCN) is applied for pixel-wise prediction in this algorithm [12]. To estimate the modelling, a stacked U-Net intermediate is used. Angles between associated edge pixels are used to classify the line segments. A convolutional neural network can be used to segment images using the eigenvector corresponding to the small eigenvalue associated with an edge pixel [13]. At the processing stage, the Hough transform is applied, and at the final stage, rotation algorithms are implemented [14] [15]. Run-length smoothing algorithm (RLSA) is applied to classify the pseudo-lines and pseudo-words extracted [16]. The ESLD (Enhanced Supervised Learning Distance) algorithm is used to evaluate the space between textual content strains, and G Clustering aids in phrase grouping or the Connected Components [17]. U-Net architecture is based on a convolutional neural network (CNN). Text Line Segmentation Using an Adaptive U-Net Architecture [18] [21].

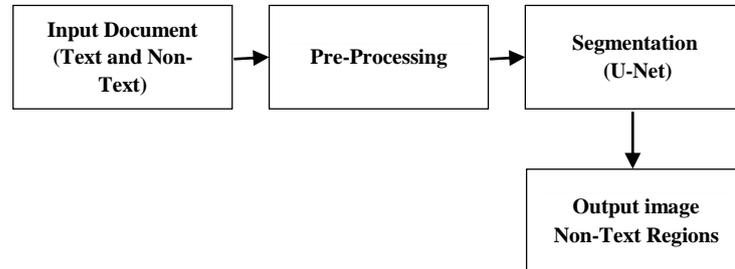
Convolutional Neural Network End-to-End Learning for Handwritten Text Segmentation Techniques. Text line segmentation are including Projection profiles, Run Length Smearing, Median segmentation, and the Bounding box method. Image Segmentation Using Mathematical Morphological Operators. Image dilation and Region labelling method by training a convolutional neural network (CNN) for evaluation of skew angle in a binary document image. The scanned document will not be skewed during the optical scanning of a document using any of the Optical Character Recognition (OCR) devices. Deep learning approach for detecting the angle of skewed scanned documents. The author has proposed using the Gabor, Wavelet, and Hough Techniques [22] [23] [24] [25]. The text lines are detected using a randomised Hough transform, and the baselines are extracted using a y-intercept histogram [26]. A method for detecting and segmenting Using a u-net deep learning network, segmentation for text-based images. In this paper, author has presented a method for estimating skew in handwritten multilingual document images in a binarization, filtration, and segmentation. A horizontal and vertical projection algorithm based on the nil aback Sauvola's and Wolf's binarization and the linear Otsu thresholding. The statistical analysis of the connected components' height. [27] [28]. Document image processing has grown in importance

in the automation of office documentation tasks Laplacian and Sobel operations for enhancing low contrast pixels in images [29]. A heat map network that converts the bounding box of a document image as well as a filtering network Gregory. [30] Text segmentation from image documents using the Symlet wavelet and 2-mean classification. [31]. Text/Non-Text Separation from Handwritten Document Images Using Back propagation Neural Networks Based on LBP and Support Vector Machine [33], from object detection to text detection and recognition. Conceptual text segmentation from synthetic images of full-text documents [36]. Text lines from flatbed scanner/camera-captured printed and handwritten documents can be segmented [37]. Deep Learning Image Segmentation and Document Layout Analysis Using CNN [39]. This paper investigated a various text line segmentation techniques, such as projection profiles, run length smearing, median segmentation, and bounding box methods [41]. The various approaches are based on the Hough transform and Document Image Page Segmentation as Semantic Segmentation. GridNet, new Convolutional Neural Network (CNN) architecture for semantic image segmentation. In deep learning techniques such as U-net over the last decade [42]. The tsegGAN: A Generative Adversarial Network for Segmenting Touching Non-Text Components from Text Ones in Handwriting in block based document image segmentation is used in this work. [43] [44]. The most difficult task in handwritten documents with multiple skews is identifying and correcting different blocks written with different skews. To the best of our knowledge, there is no work in the literature that estimates and corrects the skew of a multilingual document with multiple skews. As a result, we chose this work as the first step in handwritten/printed document analysis and segmentation, using five different scanned document input images for line and word segmentation separation from document images in text and non-text regions [45] [46].

3. PROPOSED METHOD

In this section, we discuss our approach in detail. Fig 1 shows a flowchart of the proposed method. The proposed method mainly consists of three different stages they are pre-processing, Data augmentation, segmentation for experimentation purpose we have considered real time bilingual printed (Kannada and English scripts document images. the input images containing both text and non-text (here we have considered signature & Logo) information. If the input image contains graphs and tables, the efficiency will be reduced because the proposed algorithm will not be rained for graph, tables. Since the input images real time documents, may be blurred, noisy and some distortions may present. Performance may undergo if we process the documents without removing these noises. As a result, in order to improve performance, we must improve the documents by the use of some pre-processing techniques. In the subsequent sections, we discuss the Pre-processing, Data augmentation and Segmentation.

Fig 1: Block diagram of proposed method



This influences improvement while training the network and the pre-processed yield would then be farmed into segmentation. Whenever a digital input image is divided into different subgroups to improve by decreasing complexity and make analysing simple and easy. Here, the deep constitutional neural network U-Net is used.

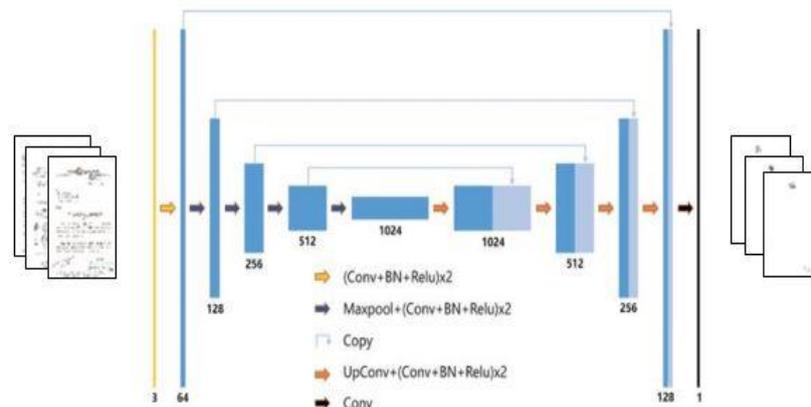


Fig 2: U-Net architecture

In this works documents all enhanced using spatial domain methods, Frequency domain methods (DFT) and Fuzzy approach, better enhancement in achieved.

3.1 Pre-processing

The real-time office document is typically scanned with a normal scanning and transformed to a jpeg image. At this point, we have data in the form of an image, which can be further analysed to retrieve the relevant information. Distraction could be present in the image obtained during the scanning process. Images could be Spattered or disrupted depending on the resolution of the scanner and the success of the technology used, such as Thresholds. Some of these disadvantages it can be eliminated by using a pre-processor, which may result in reduced detection performance eventually on. Characters that are quickly and effectively digitised.

In this works documents all improved the use of spatial domain strategies, Frequency domain strategies (DFT), and the Fuzzy approach, better enhancement in completed. As an end result, fuzzy set theory can help with a spread of uncertainties in laptop vision and photo processing programs. Fuzzy image segmentation describes a hard and fast of fuzzy image evaluation

strategies that may recognise, represent, and technique images. The three number one steps are documented image fuzzification, membership feature value change, and defuzzification. To enhance fuzzy images, gray degree mapping into a foundation features is used. The goal is to provide a greater contrasted image than the unique by means of giving more encumbrances to gray stages towards the image mean grey level than grey stages similarly from the mean.

Smoothing relates both to filling and thinning. Filling eradicates minimal breaks, gaps, and holes in digitally enhanced characters, whereas thinning reduces line width. The far more familiar smoothing method includes changing a window from around character's binary image although implementing definite standards to a content and structure of the window. So, to improve the quality of the input image, image enhancement operations such as noise removal, normalisation, binarization, and so on are performed.

Fuzzification is compelled here to map the input image with a fuzzy plane, and defuzzification is desired, i.e., the membership of a point $P_{ij}(x, y) \in D$ to the window $W_{ij}(x, y)$ are given by the equation 1.

$$W_{ij} = \frac{(P_{ij}(x,y))^\gamma}{\sum_{i=1}^n \sum_{j=1}^m (P_{ij}(x,y))^\gamma} \quad (1)$$

Where, $W_{ij}: D \rightarrow [0, 1]$

W_{ij} described the membership and $P_{ij}(x, y)$ described the pixel value. $\gamma \in (0, \infty)$ and control the fuzzification and defuzzification.

The transform ψ_{enh} is composed of the transformed W_{ij} weights multiplied by the degree of membership ψ_{ij} . Equation 2 gives the improved image.

$$\Psi_{enh}(f) = \sum_{i=1}^n \sum_{j=1}^m w_{ij} X \Psi_{ij}(f) \quad (2)$$

Where, $\psi_{ij}(f)$ represents the image (f) before enhancement. $\psi_{ij}(f)$ represents the image (f) after enhancement.

3.2 Data augmentation

In facts evaluation, information augmentation methods are used to increase the amount of records via including barely changed duplicates of pre-present facts or newly created synthetic facts from pre-current statistics. When training a machine learning model, it acts as a regularization term and helps reduce fitting problem. Minor changes to data or the use of deep learning methods to yield statistical models are applications of data augmentation. Data augmentation strategies can be an excellent tool in dealing with the complex conditions that the synthetic intelligence international faces.

Data augmentation methods could indeed improve machine learning algorithm by appears to be affected that the model can encounter in the real world. Whenever the repository for the model is rich & sufficient, the model is better and more effectively. Fig.3 illustrates the principle of the Data augmentation model.

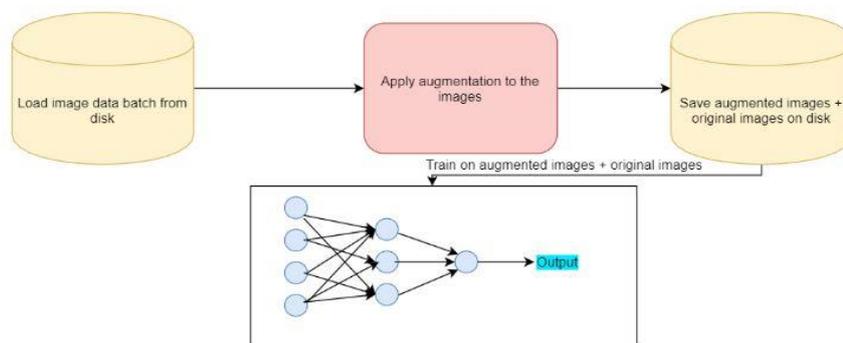


Fig 3: Training a deep neural network on both augmented

With a rich and sufficient dataset, the model is better able to estimate results more accurately and efficiently. Fig. 3 shows the working principle of Data augmentation model.

3.3 Segmentation

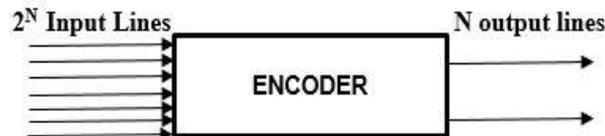
U-Net segmentation uses document images in a CNN (convolutional neural network) structure for picture segmentation. This is correct and green. Data have outperformed an earlier satisfactory approach (a sliding-window convolutional network) for segmenting axonal frameworks in electron microscopic layers. U-Net is a segmentation structure. It contains a contracting course and an expanding focus. CNN's are used to design the contracting path. The convolutions are made up of 3x3 (unpadded convolutions) and can be repeated with rectified linear units (ReLU), along with a 2x2 pooling operation with stride 2. Each down sampling step doubles the number of function channels. In the course design, convolution kernels are used. The convolutions (unpadded convolutions) are repeated and observed via rectified linear units (ReLUs) with a 2x2 maximum pooling operation for down sampling. Each step of down sampling multiplies a number of channels examined. Expansive route starts with an up sampling of the feature space, then a convolution ("up-convolution") that cuts the range of characteristic channels in half, a concatenation with the consequently cropped function map from the contractual route, and two 3x3 convolutions with ReLUs. Convolutional networks lack boundary pixels, so cropping of the image is necessary. Finally, every 64-component feature vector is mapped to an apparent magnificence label using a 1x1 convolution. It appears that this state contains 23 convolutional layers.

The input image is passed through the model by a convolutional layer with a ReLU activation function. In this case, we can see a decrease in image size from 2480X3508 to 1242X1754. Due to the use of unpadded convolutions to define the convolution layer as valid, the overall dimensionality was reduced. Additionally, there are encoder and decoder blocks on the left and right of the Convolution blocks. Fig. 2 shows architecture with encoders and decoders.

3.3.1 Encoder path:

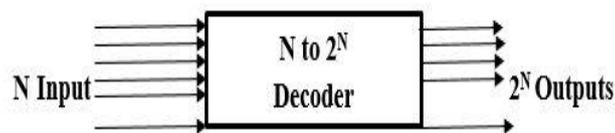
Convolution layers are composed of 3x3 kernels, 2x2 Maxpool layers, and RELU activation functions. Consequently, this reduces the feature map's dimensionality, allowing hidden layers to remain and not just the most significant ones. Connections between U-nets are presented at

the best layers, which reduces the wide range of parameters. By converting volatile statistics signals into coded messages, or analog warnings into virtual indicators, an encoder converts statistics signals into coded messages. An N bit code is represented by the N output lines resulting from the conversion of binary information into 2^N input traces. The encoder converts a signal into coded binary output when it receives an input signal.



3.3.2 Decoder Path:

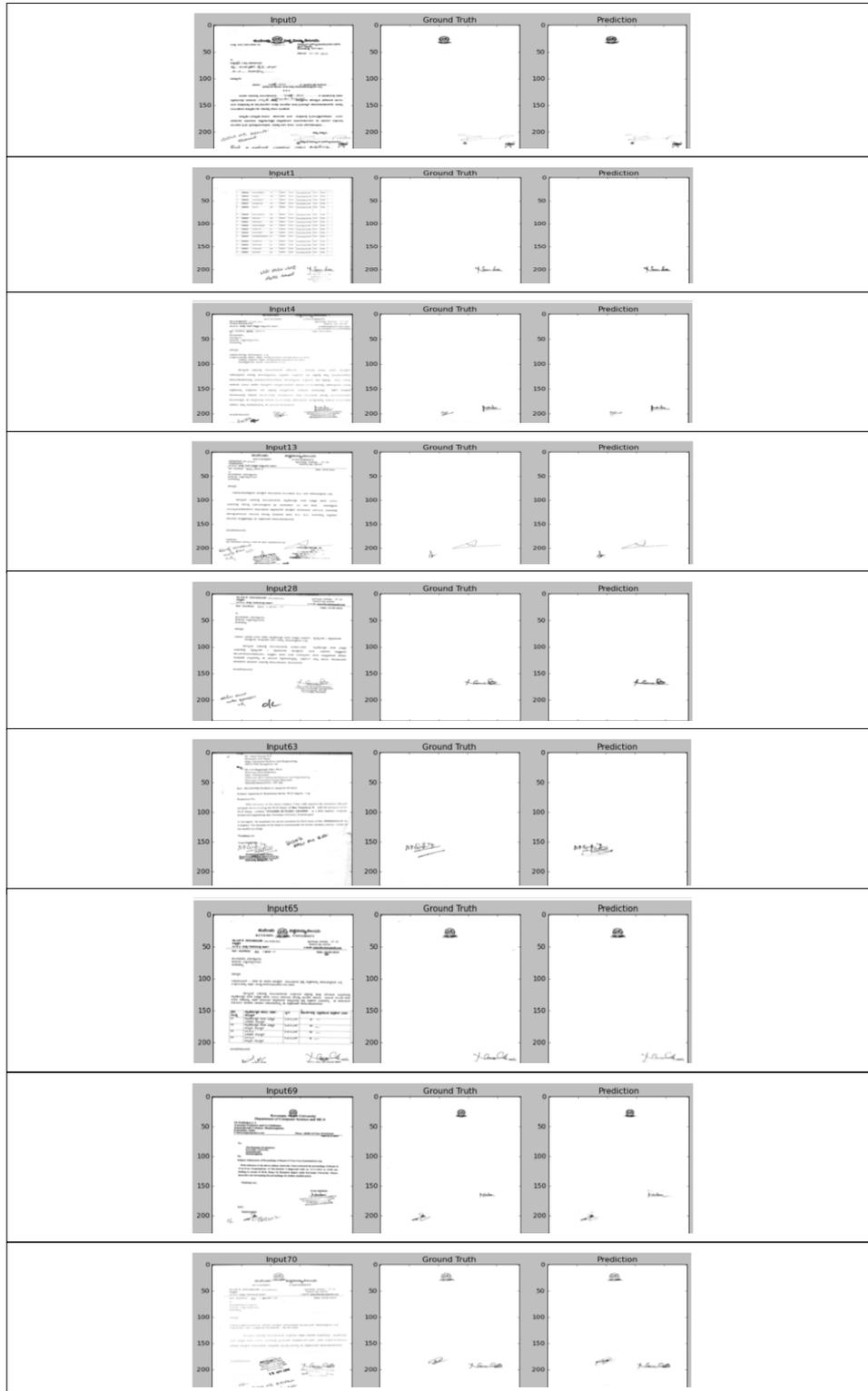
A segmented mask is made out of the input image after characteristic extraction in the encoding path. Decoders and encoders switch pooling indices during this step. The characteristic maps similar to the encoding course are copied to the decoding course. A decoder is a combinational circuit similar to an encoder; however, it operates in the opposite section. A decoder is a device that converts n traces of entering into $2n$ lines of output and generates the distinct sign as output from the coded input sign. Despite the excessive output produced by means of an AND gate, the primary interpreting element is that it produces a high output if all inputs are immoderate.



4. RESULTS AND DISCUSSIONS

The U-Net method is used to separate the non-text from the document image. The obtained output is compared to the ground truth of the respective office document images to determine segmentation outputs. We imported the Unet model ResNet as the backbone network and loaded the image mesh weights. The output is passed to the U-Net model after it describes the layout of the input intended by the base model and indeed the specially designed overlay that obtains its base mode input. The UNet model's output is then propagated to other predefined ReLU-enabled ConvNet layers. The final result is reshaped to 1242x1754. Finally, we used the base model to construct a design that takes an input (x_{inp}) and outputs an output (x_{out}).

We defined the metrics, losses, and optimizer functions after compiling the model and defining everything that fitted the training and validation data to the proposed model. After saving the model, I used the trained model to create and save the X_{train} and X_{test} predictions. After making the predictions, we defined a function that visualises the model's predictions. This function expects input and output arrays as well as predictions.



a)

(b)

(c)

We obtained a mask for same dataset of the selected training sample by randomly selecting images from the training data and defining k as zero. Then I set the figure's size and plotted all three aspects: the image, the mask, and the predictive mask. The proposed approach yielded the following results, with the ground truth and predicted output for ideal office document images shown in figure 3.

Fig. 4 shows the resultant output images 1-9, as well as the corresponding ground truth images and predicted outputs. (a) Represents input images, (b) Represents ground truth images and (c) Represents predicted images

Following training, the performance of a machine learning classifier is evaluated using key performance metrics. The confusion matrix, which is a table showing whether a classifiers needs to perform if some truth values/interests are gained, is among the performance metrics.

The most common matrices used for evaluating this architecture are accuracy, (F1) score, Precision, Sensitivity and Specificity. The proportion of classified instances pixels in an image.

$$\text{F1 Score, } F1 = 2 * \frac{P_C * R_C}{P_C + R_C} \quad (3)$$

$$\text{Precision, } P_c = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (4)$$

$$\text{Sensitivity (Recall), } R_c = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (5)$$

$$\text{Specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positives}} \quad (6)$$

Table 1: Segmentation ROC FPR and TPR

FPR	TPR
0.1	0.2
0.2	0.5
0.3	0.9
0.4	1.0
0.5	1.0
0.6	1.0
0.7	1.0
0.8	1.0
0.9	1.0
1.0	1.0

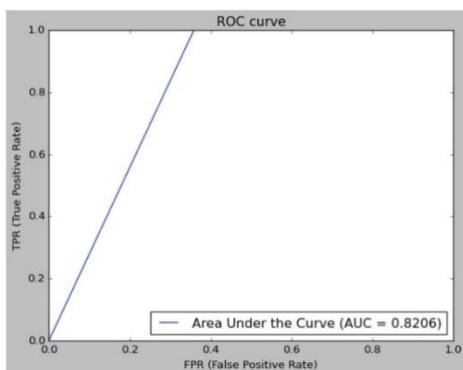


Fig 5: ROC diagrams of the proposed U-Net for Real-time dataset

The values of FPR and TPR for segmentation are given in table.1, and the ROC diagram of the proposed method is plotted in figure4. Precision and Recall diagram for the proposed method is plotted in figure 5. The accuracy diagram of the proposed method is plotted in fig 6, and the final diagram is plotted in fig 7.

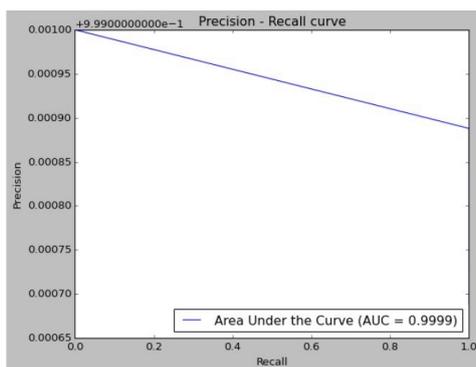


Fig 6: Precision and Recall diagram of the present work U-net Segmentation

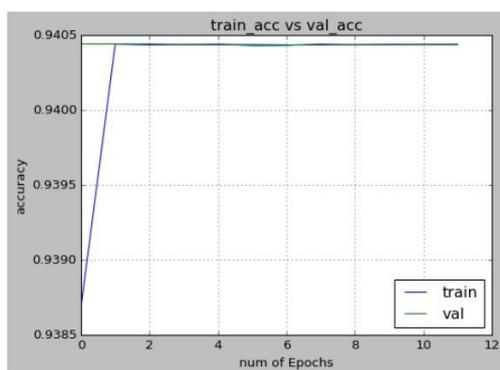


Fig 7: Accuracy diagram for the proposed method

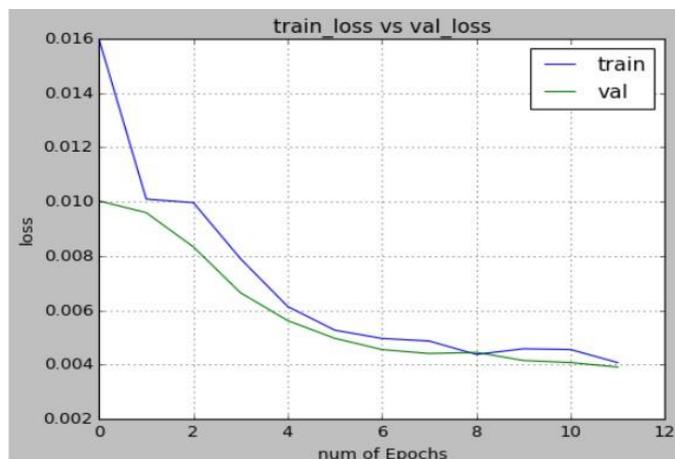


Fig 8: Loss diagram for the proposed method

Various segmentation accuracy methods are compared with the conventional method, and the results are presented in a table with a graphical representation in Fig. 8.

Table 2: Segmentation Accuracy of different methods

Segmentation Methods	Accuracy
Block segmentation	89
Watershed	94
U-Net (Proposed)	99

Fig 9: A graphical representation of the accuracy comparison of various methods.

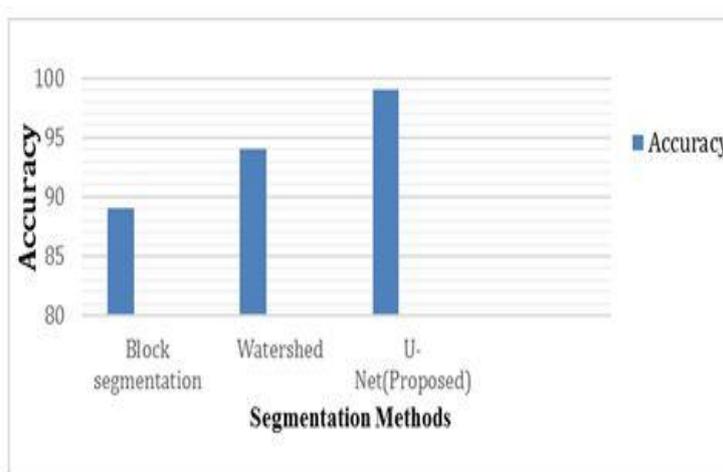


Table 3: Segmentation Accuracy of different methods

Segmentation Methods	Accuracy	Specificity	Sensitivity	Precision	F1-Score
Block segmentation	89	58	87	87	87
Watershed	94	60	90	90	90
U-Net (Proposed)	99	64	99	99	99

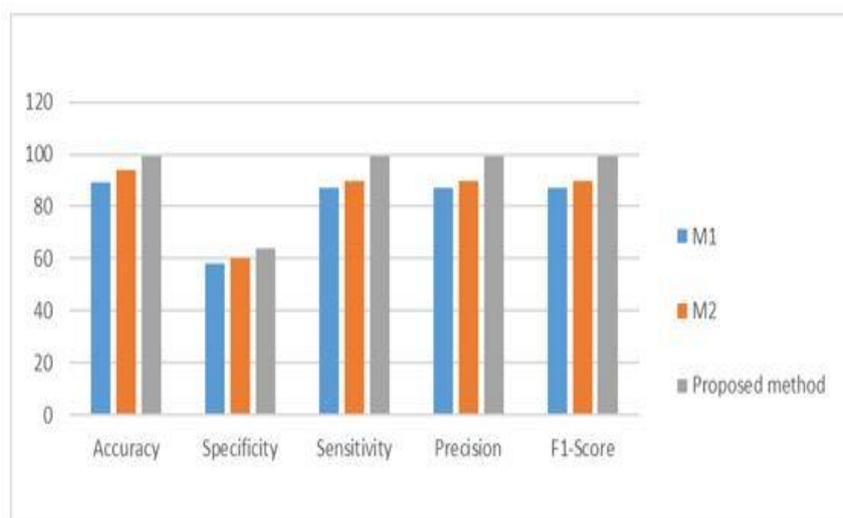


Fig 10: A graphical representation of the performance comparison of various methods

5. CONCLUSION

We proposed a deep learning approach, U-net architecture, in this paper to segment non-textual data from bilingual office document images. Experimental work is accomplished out on our own dataset, and the results show that effectiveness of the proposed technique from the results we obtained accuracy of 99%, specificity of 64%, sensitivity of 99%, and precision of 99% and F1-Score of 99%.

REFERENCES

- [1] X. Qi, Ma, L., Sun, C., J. Liu, Fast skew angle detection algorithm for scanned document images. Third Pacific-Asia Conference on Circuits, Communications and System (PACCS), 2011, pp.1-4.
- [2] K. Huang, Zixuan Chen., Min Yu., Xiaolang Yan. And Aiguo Yin, An efficient document skew detection method using probability model and q test, Electronics 9, 2019, pp.1-17.
- [3] P. V Bezmaternykh and D. P. Nikolaev, A document skew detection method using fast hough transform, in: International Conference on Machine Vision, 2020, pp.01-06.
- [4] Neha: Language independent robust skew detection and correction technique for document images, 2012, pp.111-115.
- [5] R. Srivastva, A. Raj, T. Patnaik and B. Kumar, A survey on techniques of separation of machine printed text and handwritten text, Vol. 2 Issue-3, 2013, pp.552-555.

- [6] I. V. Konya, S. Eickeler, S and C. Seibert, Fast seamless skew and orientation detection in document images, 20th International Conference on Pattern Recognition, 2010, pp.1924–1928.
- [7] A. Sakila, and D. S. Vijayarani, Skew detection and correction in the document image, International Journal of Innovative Research in Science Engineering and Technology 2017, pp.17457- 17465.
- [8] R. Pramanik and S. Bag, A novel skew correction methodology for handwritten words in multilingual multi-oriented documents, Multim. Tools Appl. 80, 2021, pp.27323–27342.
- [9] S. S. M. N. Akhter, P. P. Rege, improving skew detection and correction in different document images using a deep learning approach, 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp.1–6.
- [10] R. D. Salagar, and P. B. Patil, Application of rlsa for skew detection and correction in kannada text images’, Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp.785–788.
- [11] S.A. Oliveira, B. Seguin and F. Kaplan, dhsegment: A generic deep-learning approach for document segmentation, 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp.2018, 7–12.
- [12] Ma. K., Shu, Z., Bai, X., Wang, J., Samaras, D.: ‘Docunet: Document image unwarping via a stacked u-net’, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp.4700–4709.
- [13] Watts, N., Rani, J.: ‘Performance evaluation of improved skew detection and correction using fft and median filtering’, International Journal of Computer Applications, 2014, pp.7–16.
- [14] Bafjaish, S. S., Azmi, M. S., Al-Mhiqani, M. N., Radzid, A. R., H. B. Mahdin, Skew detection and correction of mushaf al-quran script using hough transform, International Journal of Advanced Computer Science and Applications, Vol. 9, No. 8, 2018, pp.402-409.
- [15] G. Mandal, An unprecedented approach of skew detection and correction for online bengali handwritten words, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 7, Issue 2, 2018, pp.155-159.
- [16] A Boukharouba, A new algorithm for skew correction and baseline detection based on the randomized hough transform, J. King Saud Univ. Comput. Inf. Sci. 29, 2017, pp.29–38.
- [17] B. S. Shakunthala and N.B.S. Naveen, Unconstrained handwritten kannada documents leading to line and word segmentation, International journal of engineering research and technology, Volume 5, Issue 20, 2017, pp.01-05.
- [18] A. Al-Khatatneh, S. A. Pitchay and M. Al-qudah, a review of skew detection techniques for document, 17th UKSim-AMSS International Conference on Modelling and Simulation (UKSim), 2015, pp. 316-321. IEEE.
- [19] B. Shakunthala. and C. Pillai, Enhanced text line segmentation and skew estimation for handwritten kannada document, Journal of Theoretical and Applied Information Technology 99 (1) ,2021, pp.196–206.
- [20] N.R. Soora and P.S. Deshpande, A novel local skew correction and segmentation approach for printed multilingual Indian documents, Alexandria engineering journal, 57(3), 2018, pp.1609-1618.
- [21] P.V. Bezmaternykh, D.A. Ilin and D.P. Nikolaev, U-net-bin: hacking the document image binarization contest, Computer Optics 43 (5), 2019, pp.825–832.
- [22] O. Mechi, M. Mehri, R. Ingold and N.E.B Amara, Text line segmentation in historical document images using an adaptive U-Net architecture, In 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019, pp. 369-374.

- [23] Saiyed Umer, Ranjan Mondal, Hari Mohan Pandey, Ranjeet Kumar Rout, Deep features based convolutional neural network model for text and non-text region segmentation from document images, *Applied Soft Computing*, Volume 113, 2021.
- [24] A.A. Gurav and J. Manisha, Nene, Word segmentation in document images using deep convolutional encoder decoder network, *IEEE 5th International Conference for Convergence in Technology (I2CT)*, 2019, pp.1–6.
- [25] P. Shivakumara, G.H. Kumar, D.S. Guru, and P. Nagabhushan, A novel technique for estimation of skew in binary text document images based on linear regression analysis, *Sadhana*, 30(1), 2005, pp. 69-85.
- [26] J. Jo, H.I. Koo, J. W. Soh, and N.I. Cho, Handwritten text segmentation via end-to-end learning of convolutional neural networks, *Multimedia Tools and Applications*, 2020, pp.1–14.
- [27] D.S. Guru, M. Suhil, M. Ravikumar and S. Manjunath, Small eigenvalue based skew estimation of handwritten devanagari words, in: *MIKE*, 2015, pp.216-225.
- [28] X. Zhao, Y. Yuan, M. Song, Y. Ding, F. Lin, D. Liang, and D. Zhang, Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging, *Sensors (Basel, Switzerland)* 19 ,2019, pp.01–13.
- [29] S. Deivalakshmi, P. Palanisamy and G. Vishwanathan, A novel method for text and non-text segmentation in document images, *International Conference on Communication and Signal Processing. IEEE.* 2013. pp. 255-259.
- [30] A. Gautam, Segmentation of text from image document. *International journal of computer science and information technologies*, 4(3), 2013, pp.538-540.
- [31] H. Wang, C. Pan, X. Guo, C. Ji and K. Deng, From object detection to text detection and recognition: A brief evolution history of optical character recognition, *Wiley Interdisciplinary Reviews: Computational Statistics* 13 ,2021, pp.01–32.
- [32] Z. Ibrahim, D. Isa, and R. Rajkumar, Text and non-text segmentation and classification from document images, *International Conference on Computer Science and Software Engineering*, Vol. 1, IEEE., 2008, pp. 973-976.
- [33] J. Chen, H. Shao and C. Hu, Image segmentation based on mathematical morphological operator, 2018, pp.23–41.
- [34] K.V. Jobin, and C. V. Jawahar, Document image segmentation using deep features, in: *NCVPRIPG*, 2017, pp.01–11.
- [35] L. Bures, I. Gruber, P. Neduchal, M. Hlaváček and M. Hruš, Semantic text segmentation from synthetic images of full-text documents, 2019, pp.1380– 1405.
- [36] A. Dutta, A. Garai, S. Biswas and A.K. Das, Segmentation of text lines using multi-scale cnn from warped printed and handwritten document images, *Int. J. Document Anal.Recognit.* 24, 2021, pp.299–313.
- [37] F. Lombardi and S. Marinai, Deep learning for historical document analysis and recognition—a survey, *Journal of Imaging* 6, 2020, pp.01–30.
- [38] S. Minaee, Y. Boykov, F. M. Porikli, A.J. Plaza, N. Kehtarnavaz and D. Terzopoulos, Image segmentation using deep learning: A survey, *IEEE transactions on pattern analysis and machine intelligence*, 2021, pp.01–23.
- [39] M.P. Viana, and D.A.B. Oliveira, Fast cnn-based document layout analysis, *IEEE International Conference on Computer Vision Workshops (ICCVW)* ,2017, pp.1173–1180.
- [40] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau and C. Wolf, Residual conv-deconv grid network for semantic segmentation, *ArXiv abs/1707.07958*, 2017, pp.01–12.

- [41] H. Chethana, and H. Mamatha, Comparative study of text line segmentation on handwritten kannada documents, *International Journal of Computer Science and Information Technologies* 7 (1), 2016, pp.26–33.
- [42] N. Shobha Rani and T. Vasudev, A block level segmentation of touching and overlapping characters in telugu printed and handwritten documents using iterative split analysis technique, *International journal of machine intelligence*, Bio-info publications, Volume 6, Issue 2, 2015, pp.-458-465.
- [43] R. Mondal, S. Bhowmik and R. Sarkar, tsegGAN: a generative adversarial network for segmenting touching nontext components from text ones in handwriting, *IEEE Transactions on Instrumentation and Measurement*, 70, 2020, pp.1-10.
- [44] D. Sasirekha and E. Chandra, Enhanced techniques for PDF image segmentation and text extraction. *arXiv preprint arXiv:1210.0347*. 2012, pp.01-05.
- [45] V. N. Manjunath Aradhya, H.T. Basavaraju and D.S. Guru, Decade research on text detection in images/videos: a review. *Evolutionary Intelligence*, 14(2), 2021, pp.405-431.
- [46] P. Lyu, C. Yao, W. Wu, S. Yan, and X. Bai, Multi-oriented scene text detection via corner localization and region segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp.7553-7563.