KUVEMPU UNIVERSITY

TEXT EXTRACTION FROM PRINTED BILINGUAL DOCUMENT IMAGES

A Thesis
Submitted for the Degree of

DOCTOR OF PHILOSOPHY

In the Faculty of Science and Technology

By

SHIVAKUMAR G

DEPARTMENT OF P. G. STUDIES AND RESEARCH IN COMPUTER SCIENCE

KUVEMPU UNIVERSITY, JNANASAHYADRI - 577 451

SHIVAMOGGA, INDIA

December – 2022

CERTIFICATE

This is to certify that **Sri Shivakumar G** has worked under my supervision for his Ph.D. thesis entitled "**Text Extraction from Printed Bilingual Document Images**". I also certify that the work is original and has not been submitted to any other university wholly or in part for any other degree.

07/12/22

Dr. Ravikumar M
Research Supervisor
Department of P. G. Studies and Research in Computer Science
Kuvempu University, Jnanasahyadri - 577 451, Shivamogga, INDIA

DECLARATION

I hereby declare that the entire work embodied in this Doctoral thesis has been carried out by me at the Department of P. G. Studies and Research in Computer Science, Kuvempu University, Jnanasahyadri, under the supervision of **Dr. Ravikumar M**. This thesis has not been submitted in part or full for the award of any diploma or degree of this or any other University.

Shivakumar G
Research Scholar
Department of P. G. Studies and Research in Computer Science
Kuvempu University, Jnanasahyadri - 577451, Shivamogga, INDIA

DEDICATED TO

MY FAMILY, TEACHERS & FRIENDS

"Sincerity is the sum of
All moral qualities"

-Dr. B. R. Ambedkar

Acknowledgement

I have come across many people who have supported me in the completion of the present research. The valuable time spent with scholars, well-wishers, friends, and family members relieved my stress during the study and helped me a lot to complete the research. These people stand out as having made indelible imprints on my thoughts, and their help throughout the writing of this thesis was indispensable. I feel it is my duty to express my gratitude to all of them.

I deem it a privilege and good fortune to work under the diligent supervision and guidance of Dr. Ravikumar M, for his invaluable scholarly advice, constant guidance and support have always kept me going ahead in my life. He taught me what research is all about and how to be a good researcher. His passion for research remained motivational. Their commitment to research and guidance is priceless. I owe a great debt of gratitude to my supervisor.

I wish to express my deepest gratitude to the family of my supervisor Smt. Bhagyalaxmi madam, for her hospitality and for treating me like one of their family members. I could not forget the wonderful memories and occasions that I have spent with their family and children of my guide, Bharath R and Hamsika R.

I am indebted to my parents for providing me with the right education and for inculcating dedication and discipline in my life. The love and affection bestowed on me for achieving this goal cannot be expressed in words. I am grateful to my father Sri. Guruswamy. T who supported me in all my endeavors and encouraged me throughout my research journey. I thank my mother Smt. Indiramma for her endless love, affection, and blessings in every walk of my life. The love and affection of my parents gave me the strength to complete this work. I take this opportunity to express my gratitude to my in-laws' Sri. Sathyanarayanswamy Nayaka and Smt. Jayalakshmi for their blessings and encouragement in completing my research work.

My special thanks to my beloved sister Komala G, brother-in-law SandeepNayaka G. S, I cannot forget the happy moments that I spent with my niece Namsvi M.S. and Namratha M.S. Their affectionate and endless love.

My heartfelt thanks to grandfather Sri. Thippeswamy., and grandmother Smt.Sathyamma, grandfather Sri. Obaiah K, grandmother Smt.Palamma, and brother-in-law Srinivasa, Kanumaiah, for their concern and blessings. My special gratitude to my uncle Kumar Swamy K O, and Mamatha L, for helping me throughout my research period. Niece Monika, nephew Sai Akhil, and Sai Anirudh for their love and support. I am grateful to all my family members.

I thank my uncle Obaiah P., and aunt Yashodhamma., for the love and affection they bestowed on me. I thank my sisters Anu, Thanu, Jayashree, Dhanalakshmi (chintu), Aishwarya, Nikihla, Sowmya, and brothers Boresh, Dhananjaya, Raki and Suri, nephews Shashi and Mohan for their love and support in the completion of the present research.

My sincere thanks to Dr. Suresha M., Dr. Prabhakar C J., and Dr. Yogish Naik G R., for their suggestions and critical remarks which have helped me to improve the quality of my research. I also thank all the teaching and non-teaching staff of the Department of Computer Science, Kuvempu University, Jnanasahyadri, and Shankaraghatta for their constant support in the department.

I am grateful to Dr.D.S. Guru sir for his invaluable guidance and regular encouragement throughout my research. His assistance to complete my research is valuable.

I am grateful to Dr. Ashok Rao, Dr. S. Manjunath, Dr. Vinay Kumar N, Dr. Imran, Dr. Mahamad Suhil, and Santhosh Kumar K S, for their dynamic guidance.

My sincere thanks to Narasimha Murthy S K, Dr. Gireesh B J, and Dr. Venkatesh, Department of Mathematics, Kuvempu University for their continuous support in the department.

I shall not forget the words of my senior researcher Dr. Chandrashekhar H, Dr. Shreekanth K N, Dr. Punith Kumar B S., and Dr. Sandeep., which have helped me come up with quality research. I also thank all my research friends Dr. Shivaprasad B.J., Sampath Kumar S, Prasanth M C, Rachana P G, Omar Ali, Raghukumar D S, Kuppa S, Divyarani R., Rabhiya., Shrinivasa S R, Vidyasagar K B., Vidyasagar S D, Dr. Nagaraj, and Narendra R for their care.

My cordial thanks to Pavana Suresh., Poornima M R., Yashas, Manaswini, Lanchan Poornachandra, and Mahanthesh for their concern, hospitality, and their valuable time.

This study would not have been possible without the cooperation and support extended by Ramesh R., Rakesh P., Nishanth H., Govindha B J., Vrushanka., Girisha., Siddaraju., Manju, Shilpa., Ayesha., Raghu., Shobha., Vani., Laxmi R., Manjula., Bharathi and It will be endless to list each and every friend and relative of mine who have directly or indirectly wished my success in this endeavor. I thank one and all.

I would also like to thank the Staff members of the Examination, SC/ST Development cell, and Academic and Finance Section of Kuvempu University for their help. I would like to acknowledge Dr. Walmiki R, Librarian, Shridhar S, Assistant Librarian, and Shankarappa B, Junior Computer Programmer, University Central Library, Kuvempu University for their support.

Finally, my thanks to all the people who have supported me directly and indirectly to complete the research work.

I acknowledge and am indebted to my all relatives, friends, and well-wishers, whom I have not mentioned above and whose best wishes have always encouraged me. Above all, I thank the Almighty God Sri Gurunathaswamy, Sri Horgelamma, and Sri Marikambadevi who shower their blessings on me and grace all walks of my life.

Shivakumar G

List of Figures

List of Tables

.

Contents

Abstract

World is now facing towards digitization, which has made it inevitable for transformation from paper to paperless office using document image processing. This will reduce time and space in processing and storage of documents. Whereas the document of the image will contain required supportive data in an image such as pictures, texts, tabular columns, etc. Document image processing has become major research area, where in document images are processed and stored in the digital form. Documents may contain text, non-text or both as a whole. There might be skew present in the document which must be corrected in order to distinguish text from non-text.

Text is an ordered sequence containing piece of writing distinct from other material such as diagrams, graphs etc., Text can be in printed or handwritten form which results in classification as printed document images, handwritten document images or both, where the document can be completely in printed form or completely in handwritten form or it might contain both. Further, documents can be classified into monolingual, bilingual, trilingual or document with different languages. Each language needs individual OCR system for its identification. So, multilingual document need multilingual OCR system for its identification and processing. Non-text is anything other than text in document, it includes maps, charts, graphs, illustrations, diagrams, photographs, expressions, formulae etc.,

The problem of text information extraction needs to be defined more precisely before processing further. Text Information Extraction (TIE) system receives on input still image and sequence of images. The images can be in gray scale or color, compressed or un-compressed and the text in images may or may not move. Text from image can be considered as human clear arrangement of char words they frame that can be encrypted computer comprehensible from American standard code information interchange (ASCII). Text is commonly well-known as of non-character encrypted documents, such as detailed images into the form of bitmaps and database, which is now and then referred to as presence in binary. In other form of text removal is method by which we transform printed text/scanned page or image in which text are available to ASCII character that a computer can identify.

Text and Non-text segmentation and classification is very important in document layout analysis system before it is presented to an OCR system. The text detection and extraction can be divided into the following steps. They are:1) detection 2) localization 3) tracking 4) extraction and enhancement 5) recognition(OCR). In this direction, a small amount of work is carried out in the Indian context. Hence, this has motivated to consider the study of handwritten Kannada vowels and English uppercase alphabets recognition system as the initial, work to meet the objective of processing bi-lingual (Kannada \& English) documents.

Chapter 1

Prologue

1.1 Preamble

World is now facing towards digitization, which has made it inevitable for transformation from paper to paperless office using document image processing. This will reduce time and space in processing and storage of documents. Whereas the document of the image will contain required supportive data in an image such as pictures, texts, tabular columns, etc. Document image processing has become major research area, where in document images are processed and stored in the digital form. Documents may contain text, non-text or both as a whole. There might be skew present in the document which must be corrected in order to distinguish text and non-text separation from document images.

Text is an ordered sequence containing piece of writing distinct from other material such as diagrams, graphs etc., Text can be in printed or handwritten form which results in classification as printed document images, handwritten document images or both, where the document can be completely in printed form or completely in handwritten form or it might contain both. Further, documents can be classified into monolingual, bilingual, trilingual or document with different languages. Each language needs individual OCR

Some parts of the materials in this chapter have appeared in the following research paper.

1. Ravikumar M., and Shivakumar G 2020. "A Survey on Text Detection from Document Images", In International Conference on Intelligent Computing and Smart Communication. Springer, Singapore, pp. 961-972. (springer).

system for its identification. So, multilingual document need multilingual OCR system for its identification and processing. Non-text is anything other than text in document, it includes maps, charts, graphs, illustrations, diagrams, photographs, expressions, formulae etc.,

The problem of text information extraction needs to be defined more precisely before processing further. Text Information Extraction (TIE) system receives on input still image and sequence of images. The images can be in gray scale or color, compressed or un-compressed and the text in images may or may not move. Text from image can be considered as human clear arrangement of char words they frame that can be encrypted computer comprehensible from American standard code information interchange (ASCII). Text is commonly well-known as of non-character encrypted documents, such as detailed images into the form of bitmaps and database, which is now and then referred to as presence in binary. In other form of text removal is method by which we transform printed text/scanned page or image in which text are available to ASCII character that a computer can identify.

Text and Non-text segmentation and classification is very important in document layout analysis system before it is presented to an OCR system. The text detection and extraction can be divided into the following steps. They are:1) detection 2) localization 3) tracking 4) extraction and enhancement 5) recognition(OCR). In this direction, a small amount of work is carried out in the Indian context. Hence, this has motivated to consider the study of handwritten Kannada vowels and English uppercase alphabets recognition system as the initial, work to meet the objective of processing bi-lingual (Kannada & English) documents.

In a wide range of application text/non-text separation plays a vital role in document analysis contents, machine-printed and handwritten texts always intermixed appear in several kinds of documents especially in office/note/... documents. A document is usually composed of two parts, one is the preprinted-machine texts and another is the handwritten texts. The recognition methodologies and mechanisms for machine-printed and handwritten texts are totally different. To achieve the optimal performance, we have to distinguish these two different types of texts. Once the text is identified as printed machine, it is sent to the printed optical character recognition kernel. Otherwise it is sent to the handwritten character recognition kernel. The identification of machine-printed and handwritten texts is thus must for later optical character recognition process. It is also a research issue in the area of document analysis (DA) and optical character recognition (OCR) non-text is anything other than text is document, it includes maps, charts, graphs, illustration, diagrams, photographs, expressions, formulae etc.

Existing text and non-text separation methods generally classified into four categories: edge-based, linked component-based, texture-based, and eigenvalue-based. The text components in high quality and basic backdrop images are extracted using connected component optimization techniques. Texture approaches identify textual information in complicated background images, but classifying text and non-text components needs extensive classifiers. Edge-based approaches identify text by considering sudden changes, rendering it faster than texture and connected component methods. In low-resolution images, eigenvalue-based approach separates the text and non-text components.

However, in the actual world, the text may be used in a horizontal or unstructured orientation. Because India is a multilingual country, the text in images may be in various languages. Since the last decade, different text identification systems for horizontal, arbitrary orientation, and bilingual languages were proposed. Each of these methodologies contributed to the research world in its method. As a result, we have tried to divide the advantages and disadvantages of horizontal text detection methods, randomly oriented text detection methods, and bilingual text and non-text detection methods. Finally, bilingual text detection is the final stage, which includes detecting various languages in a single image. Only caption text, either image text, or both at once are present in these three levels. The term caption text refers to artificial or superimposed text, whereas image representation refers to writing that exists naturally. Where orientation-based approaches are provided by horizontal and arbitrary oriented categories, and the textual aspects for diverse geometrical forms text are presented by the bilingual stage. The combination of orientation methods and text feature approaches yields a method for text detection that is both rapid and reliable. As a result, this research contributes something new to the current survey publications. This new category of text recognition methodologies assists forthcoming scholars in understanding from pre-processing methods, major methods, post-processing methods, as well as measuring terms used in a certain content.

1.2 Document Image Processing in Machine Learning

The recognition of printed scanned document images, also known as Optical Character Recognition (OCR), was one of the first applications of Document Image Analysis(DIA). Computer vision is a part of Artificial intelligence that includes mechanisms for acquiring, processing, analyzing, and able to understand document images. The document images are obtained from real-world applications for the purpose of generating both symbolic and numeric data. The primary goal of this research is to duplicate human vision capabilities by electrically perceiving and able to understand images. Computer vision technology typically requires a combination of low-level image processing to improve image quality and high-level pattern recognition and image recognizing to identify features in the document image.

A process like this is made up of several steps, including removing scanner noise, identifying sets of field labels and field values, and finally recognizing the text. All of these steps are difficult, and considerable research effort has been expanded to solve these issues.

There are many more topics that are the subject of active research.The typical document image processing pipeline is composed of 3 generic steps.

1. Pre-processing - Noise removal, blur removal, rectification, deskewing, binarization.

2. Layout Analysis - Understanding document structure to, e.g., identify Regions of Interest (RoI).

3. Recognition - Extracting application-specific information from each RoI corresponding to bounding box.

Pre-processing and layout analysis stages are generally dependent on the types and structure of the document images in consideration. Segmenting form images, for example, is not the same as segmenting text lines in a printed/handwritten document image. The end application determines the detection and recognition process. The goal of most apps is to transcribe textual and non-textual content.

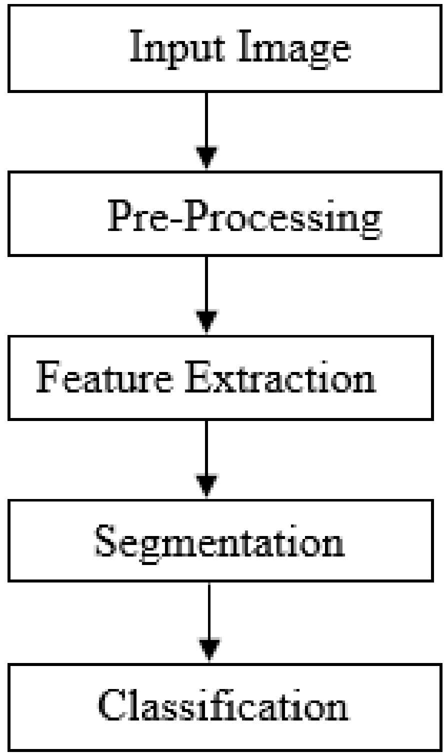


Figure 1.1: Steps involved in text and non-text separation using machine learning

The machine learning system's image acquisition can be defined as the system that receives an image from a source, which is usually hardware based. In most machine learning systems, this is the first stage. There can be no processing without an image. Besides from the type of modes, the digital image is obtained by one or more document images. It also includes business letters, bank cheques, challans, circulars, doctor's prescriptions, postal department documents, engineering drawings and maps, test halltickets, incomtax letters, invoices, petitions, requests, purchase bills, symbolic data, technical manuals, and more. All of these document images are subjected to digital image processing techniques in order to extract, retrieve, modify, transmit, and reuse information. Pre-processing is an operation that is performed at the lowest level of abstraction on input and output images. The goal of pre-processing is to improve some image attributes that are relevant for later processing as well as to enhance distorted and degraded image data. The brightness transformation, geometric transformation, local neighbourhood pre-processing, and picture restoration are the most common pre-processing approaches. The image data can be used to extract the various features needed for processing. Lines, edges, ridges, corners, points, texture, colour, and shape are examples of some of the features. The goal of image segmentation is to separate an image into a number of segments that have a strong relationship with real-world items or places. The methods of segmentation are divided into three categories: threshold segmentation, region-based segmentation, and edge-based segmentation.

Data ambiguity is one of the most common segmentation issues, which is often increased by information noise. The result of segmentation can be improved by using more and priori information from the image. This stage usually requires a small set of information as input. This stage involves analyzing if the data terms to refer the model's and the application's assumptions. To determine application specific parameters such as object postures and size. To classify the detected object into various groups. In medical, military, security, and pattern recognition applications, the final decisions required are to pass or fail automatic inspection applications, to match or no match in recognition applications, and to flag for further human study. A subset of machine learning is pattern recognition. Machine learning is based on the detection of patterns and regularities in data. It is also defined as the process of categorizing input data into patterns based on key features. In most cases, pattern recognition is combined with machine learning. In general, pattern recognition techniques formalise and explain the visualised pattern. However, machine learning focuses on increasing the recognition rate. This method usually yields better results.

Image processing, image acquisition, and image analysis are fundamental components of machine learning. The pattern recognition technique is important in the machine learning system because it allows for image analysis as well as recognition. The first step is to acquire an image. Image acquisition is accomplished through the use of digital imaging techniques, which are then processed by computers. The image segmentation process is the next step. Image segmentation is the process of extracting a specific region of interest from an image. During this process, the object is separated from the other regions as well as the background of the image. Noise reduction and image sharpening can also be performed during segmentation. The third step is the feature extraction process. The goal of feature extraction is to classify an object using a measurement value. Objects with similar measurement values are classified into one category, while objects with different measurement values are classified into a different category. In this step, specific image features are extracted. These features are referred to as feature vectors, and they are useful for describing the image's content. The fourth step is classification. During this step, the objects are classified into various categories. If predefined categories are used, it is referred to as supervised classification; otherwise, it is referred to as unsupervised classification. Various validation methods can be used to estimate the resultant classifications during the post-processing process.

1.3 Classification of Document Image

Image classification is a current research topic. Image processing pattern recognition techniques are now used for real-world objects such as medical images, document images, satellite images, and so on. The focus of this research is on segmentation, feature selection, and classification. When images are classified, especially text and non-text document images, a validation problem arises. Different features should be chosen for each category to differentiate them. As a result, the feature can depict the object's required properties. Descriptors derived from images are used in the image classification process. Colors, textures, and shapes are all important image characteristics that are used in image classification. These image characteristics are specified by computing various types of descriptors.

A number of different of descriptors are used to describe the content of an image. For identification and classification, the proposed method considers circular shaped seals, logos, signatures, and Kannada/English uppercase and lowercase characters. When compared to machine printed text, which is uniform in nature, identifying handwritten text is a critical task. The reference processing of complex documents is a difficult task for live retrieval and recognition document applications. Algorithms that work well on simple documents may not work well on complex documents (which contain a mixture of noise, handwriting, machine printed text with different fonts and font sizes, seals, tables, stamps, and rule lines) because the elements impose many constraints on the algorithms.

The primary task of processing such document images is to isolate the document's various components. Documents that have been separated into components are known as indexed documents. Document indexing can be done using textual and/or graphical entities. Organizations are currently implementing digital document images to improve the efficiency of paper-intensive workflows and to reduce the workload of processing information from office document images, faxes, invoices, reports, and so on. Organizations are compelled by this digital document images to automate the distribution of incoming mails to their respective departments based on the content of the digital documents. For verification and validation of the proposed algorithm, official forms with machine printed text in various font sizes and handwriting patterns are used. It also includes some graphical elements such as a seal, logos, signatures, and sign images.

1.4 Applications of Document image processing using Machine learning

With the growing variety of tasks and domains, the fundamental techniques used in DIA have shifted. Early applications relied heavily on heuristic rules that applied to specific document domains but did not generalize well to all document classes. Supervised machine learning techniques are now widely used, in which a generalizable model is learned from human-annotated examples of input and output pairs. Rather than developing new heuristic rules for each language/domain/task combination, the same machine learning model can be applied to different sets of labeled data. While traditional machine learning models operate on user selected task-specific features, a class of hierarchical models known as deep neural networks learn superior task-specific features directly from image pixels.

Deep learning has recently transformed many fields of research, including computer vision, natural language processing, and bioinformatics, and DIA is no exception. Each layer of the deep network abstracts the previous layer's representation, gradually transforming an input image into high level semantic information such as document type. A document is typically a piece of paper that contains printed and handwritten text as well as tables, graphs, stamps, seals, logos, circuit diagrams, pictures, and so on. A document can be simple or complex. A wide range of information that was previously stored on paper is now converted into digital form using a scanner or fax machine for efficient storage and intelligent processing. Common documents include college admission application forms, business letters, bank cheques, challans, circulars, doctor's prescriptions, postal department documents, engineering drawings and maps, exam halltickets, incomtax letters, invoices, petitions, requests, purchase bills, symbolic data, technical manuals, and so on. All of these document images are processed using digital image processing techniques in order to extract, retrieve, modify, transmit, and reuse information.

1.5 Literature survey

From the literature, we found that most of the works on text extraction is monolingual and bilingual documents various types, in that it is oblivious for containing the bi-lingual scripts those are a regional script (Kannada) and International script English(Roman). It is the combination representation of language each document has got its own characteristics. There have been several techniques proposed for in the literature for text extraction. Text detection and extraction can be achieved using the feature extraction methods. Various preprocessing techniques are available in the literature, some of the techniques are highlighted, they are Mathematical Morphological operations and Dilation (Adesh Kumar et al., 2015; Anuj Singh et al., 2014; xiao-weizhang et al., 2008; ChaudhariShailesh A. et al., 2015; Dhandra B.V. et al., 2012), image binarization, thresh-holding approach (Ankit Kumar et al., 2012),Gabor Filter (Kumary R Soumya et al., 2014; PeetaBasaPati et al., 2004; Anubhav Kumar et al., 2014), Gray-scale transformation, Smoothing, Contrast Enhancement and BBs (Jian Yuan et al., 2009; Paraag Agrawal et al., 2012; Sourav Ghosh, et al., 2018), Image edge detection, Sobel, Prewitt, Laplacian of Gaussian (Zhihu Huang et al., 2014; Shekar B.H. et al., 2015). Re-sampling, Geometrical properties and Otsu algorithm (DanialMd Nor et al., 2011; Hye-Ran Byun et al., 2002; Frank D. Julca-Aguilar et al., 2017; Lei Sun et al., 2015), deblurring methods & kernel estimation (Hojin Cho et al., 2012).

Contrast edge detection using rough set theory (Radhika Patel., et al 2015), SWT and canny edge detector, (Najwa-Maria Chidiac et.al., 2016), TIE&Text extraction algorithm(Chowdhury S P et.al., 2009; Mamatha B S et al., 2014; ShervinMinaee et al., 2017; Uma B. Karanje et al., 2014; Karanjeet al., 2009), Block Based local thresholding (Yassin M. Y. et al., 2000), MSER & SWT (Hui Wu, et al., 2016), Text detection (Uma B. Karanje et al., 2014), heterogeneous parallelization scheme (Yun Song et al., 2017), Document Image Analysis (DIA) (Vikas Yadav et al., 2016). In the area of segmentation, major techniques used are specified, such asHaar and Discrete wavelet transform (DWT) (Adesh Kumar et.al., 2015), The Horizontal projection (Ankit Kumar et.al., 2012), A two-dimensional wavelet transform and K-means clustering algorithm (Anuj Singh et.al., 2014; xiao-weizhang et.al., 2008), Gabor filter, Morphological operation and Heuristic Filtering process, (Anubhav Kumar, 2014), Color chain segmentation. (Chowdhury S P et.al., 2009), Otsu method & Geometric (DanialMd Nor et.al., 2011), connected components (Frank D. Julca-Aguilar et.al., 2017; Viet Phuong Le et.al., 2015), Bottom-UpApproach, Clustering, Top-Down&MultiscaleStrategy, HoughTransform technique, morphological postprocessing& x-projection techniques (Kumary R Soumya et.al., 2014),Deblurring methods & kernel estimation, (Hojin Cho et.al., 2012), Connected components extraction and non-text filtering (Hui Wu et.al., 2016),8 – connected pixel connectivityand NN matrices(Paraag Agrawal et.al., 2012), Texture based segmentation algorithm (PeetaBasaPati et.al., 2004),Text segmentation, k-means &sparse decomposition (ShervinMinaee et.al., 2017), Prewitt edge detection algorithm, Canny edge detector (Jian Yuan, et al., 2009).

FAST (Features from Accelerated Segment Test), (Vikas Yadav et al., 2016), Markov Random Field (MRF) and a Conditional Random Field (CRF), (Uma B. Karanje et.al., 2014), Block segmentation(Zaidah Ibrahim Dino Isa et.al., 2014). Gabor function based multichannel directional filtering, (PeetaBasaPati et.al., 2004), Line removal, Discontinuity &Dot removal (Paraag Agrawal et.al., 2012), discriminating features and Binary Tree classifier (Priyanka P. Yeotikar et.al., 2013), PCA, connected component &Block blob (Radhika Patel et.al., 2015), Text localization and Kirsch Directional Masks, (Shekar B.H et.al., 2015), Edge based text extraction algorithm (ShervinMinaee et.al., 2017), Connected component, GLCM, LBP, NB, MLP, SMO, K-NN and RF (Sourav Ghosh et.al., 2018), MSER, SIFT, Connected component HOG and SVM (Uma B. Karanje et.al.,2014; Yun Song, et.al., 2017), Component and edges extraction (Vikas Yadav et.al., 2016). Those features utilize size, shape, stroke width and position information of connected components (Viet Phuong Le et.al., 2015), Morphological gradient, Nonlinear filter and CCs (Yassin M. Y. Hasan et.al., 2000), Edge detection methods (Sobel, Prewitt, Laplacian Gaussian and canny method), 8-connected objects detection algorithm &SWT (Zhihu Huang et al., 2014) Used.

To process the documents effectively, enhancement of documents is very much essential [(Raman Maini et al., 2010; Iwasokun et al., 2014; Vaquar et al., 2019), Atena Farahmand et al., 2013; Firdausy, Kartika et al., 2007; Purba, Angga et al., 2019; Z. Shi, et al., 2011;

Poonam et al., 2014; Shi, Zhixin et al., 2004; Ravinder Kaur et al., 2016; Satnam Kaur et al., 2017; Harraj et al., 2015; Sitti Rachmawati Yahya et al., 2010; S. Perumal et al., 2018; Rubina Parveenet al., 2018; Ganbold Ganchimeg 2015; Harmandeep Kaur Ranota et al., 2014; Brindha, Bharathi et al., 2015; Di Lu et al., 2018; Sitti Rachmawati Yahya et al., 2009; Reza Farrahi Moghaddam et al., 2009; Parashuram Bannigidad et al., 2016; Jianbin Xiong et al., 2021; Mustafa, Wan et al., 2018; Sugapriya.C et al., 2017; V. Magudeeswaran et al., 2013; Samrudh. et al., 2018; Sarath K et al., 2017; Puri, Shalini et al., 2020; A. Thakur et al., 2015; Zhixin Shi et al., 2004; Sattar, Farook et al., 1999; Antoni Buades et al., 2005; Ravikumar M., et al., 2020; Ravikumar M. et al., 2020; Chidiac et al., 2016; S.P. Chowdhury et al., 2009; Y.M.Y. Hasan et al., 2000;, H. Wu et al., 2016; B.H. Shekar et al., 2015; V. Yadav et al., 2016)].

In this section, we discuss the related work on text extraction/detection from both printed as well as handwritten document images.

Using different text extraction methods like region-based method, edge method, texture method, morphological-method, text from an image is extracted (K.R. Soumya et al., 2014). After the detailed survey given on comparison and performance evaluation and different text extraction methods, it is found that region and text-based methods give poor result compared with the morphological and edge-based methods.

By using edge-based and K-means clustering algorithm, text extracted from live captured image is with diversified background (A. Singh et al., 2014). K-means clustering algorithm is performed on dataset,which partitions)into group according to some distinct distance measure. Non-text region from an image is removed using morphological operations. After the experimentation the overall precision rate and recall rate result was compared with the edge-based algorithm and connected component-based algorithm.

A two-dimensional wavelet transform used for text extraction is proposed (X.-W. Zhang et al., 2008), For the purpose of classification of the images into text region, simple background region, and complex background region, k-means clustering algorithm is used. After the classification is performed, clustering is done using morphological operation. Experimentation is carried out on 100 diverse gray-scale pictures, which contain content data with distinctive languages, textual styles, and sizes.

Text detection and extraction from natural scene images, which are captured through mobile camera and digital devices is proposed (J. Yuan et al., 2009), The proposed algorithm also tackles the complications involved in scene images like uneven illumination and reflection, poor lighting conditions and complex background analysis. Sharp transitions are detected using a revised Prewitt edge detection algorithm. The image is segmented into several regions. Each region can be regarded as an object. Finally, it is considered as abnormal objects (area too large or too small, width is far longer than height, etc).

Text regions are detected by 8-connected objects in natural scene images using which region-based extraction method is proposed (Z. Huang et al., 2014). Image can be detected using the median filter which is used to reduce the noise present in the image. In order to improve the precision of edge detection methods (Sobel, Prewitt, Laplacian Gaussian and canny method), experimentation is carried out on ICDAR-2014 dataset containing 509 English images, in which 258 images are taken for training set and 251 images are taken for testing set. To improve the performance of Stroke Width Transform method (SWT), the modified method is conducted.

Some more work are addressed on Signature and Logo detection [(Romit Beed et al., 2018), (Rajesh et al., 2015), (Umesh et al., 2017), (Luiz et al., 2017), (HWEI-JEN LIN et al., 2001), (Ilkhan Cuceloglu et al., 2018; Ranju Mandal et al., 2011; Ranju Manda et al., 2013; Mohammed Javed et al., 2013), (Corinna Cortes et al., 2000; Wafa Elmannai et al., 2012; Mayada et al., 2010; Sabourin 1988; Nabin Sharma et al., 2018; Ranju Mandal et al., 2013; Sheraz Ahmed et al., 2012), (Hongye Wang et al., 2009; Zhe Li et al., 2010; Naqvi et al., 2011; Umesh et al., 2015; Sheetala et al.,2015; Umesh et al., 2013; Vaijinath et al., 2017; Nabin Sharma et al., 2018; Rifiana et al., 2018; Tuan et al., 2003; Mohammed et al., 2013; Jans et al., 2017; Showmik et al., 2018; Afsoon et al., 2017; Mohammad et al., 2012; Yifei et al., 2014; MatheelE et al., 2017; Alireza et al., 2014; Alireza et al., 2013; Smita et al., 2016), (Romit Beed et al., 2018; Rajesh. T.M et al., 2015; Umesh D. Dixit et al., 2017; Luiz G. Hafemann 2017).]

In this chapter (P.B. Pati et al., 2004), the authors have proposed a Gabor function-based multichannel directional filtering approach that is used for separation of text and non-text regions from the images (containing graphs, natural images, and other kinds of sketches drawn with lines) is proposed. Experimentation is carried out on images of 1000 words (Tamil, Hindi, Odiya, English). The documents considered for experimentation are bilingual, where English is a common script. Using linear discriminant function, script is identified for document containing Hindi and English.

Image segmentation and text extraction from natural scene images are proposed (D.M. Nor et al., 2011). Using Otsu method where Geometric Properties are used text localization and extraction are performed by using connected component algorithm and Run Length Smoothing Algorithm (RLSA) approaches. The proposed algorithm gives a reliable OCR results.

Using text-specific properties, text image deblurring is proposed for text-specific image deconvolution approaches (H. Cho et al., 2012), where the proposed algorithm not only estimates a more accurate blur kernel but also restores sharper texts. Experimentation is carried out on both blurred and also deblurred images, the Peak Signal Noise Ratio (PSNR) value is found to be 15.66 for blurred and 28.52 for deblurred images. Experimental results show that the proposed method generates higher quality results on deblurring text images.

Some good number of algorithms are attempted towards segmentation and skew estimation of non-text information In this section, In most of the cases, documents will be multilingual in nature comprising different languages with multiple skews. A multiple skew estimation technique is proposed (Guru et al., 2013), (Tang Y et al., 1996), (Kasiviswanathan et al., 2010) where skew is estimated by fitting a minimum circumscribing ellipse and k-means clustering is used to estimate skew of multiple blocks. A method for estimating document image skew angle is presented (Shah et al., 2014), (Babu et al., 2006), (Kavallieratou et al., 2002), (Singha et al., 2008) where it depends on objects with rectangular shape such as paragraphs, texts, and figures. The angle of that rectangle represents the angle of document skew. Skew detection and correction using linear regression technique is proposed (Wagdy et al., 2014), (Boukharouba, et al., 2012) and this method uses the Hough transform to detect skew with large angles.

A method for skew estimation in binary images is proposed. This method is based on binary moments, where moment-based method to each binary object evaluates their local text skews (Brodi´c et al., 2012). A geometrical technique for line and word segmentation is presented in (NarasimhaReddy et al., 2017) which also estimates multiple skews if present in the document and corrects it by natural method which helps in finding top and bottom border points of shirorekha, and accuracy of 94% is recorded for Indian government office documents. [(Shah et al., 2014), (Ravikumar et al., 2017), (Ghosh, et al., 2012), (Rani, et al., 2015), (Ramakrishna Murty et al., 2011), (Pramanik et al., 2021; Akhter et al., 2020), (Salagar et al., 2020), (Oliveira et al., 2018), (Ma et al., 2018), (Bafjaish et al., 2018; Mandal et al., 2018), (Boukharouba et al., 2017), (Shakunthala et al., 2021), (Khatatneh et al., 2015; Bezmaternykh et al., 2021), (Mechi et al., 2019; Saiyed et al., 2021; Gurav et al., 2019; Shivakumar et al., 2005), (Jo et al., 2020), (Guru et al., 2015; Zhao et al., 2019), (Deivalakshmi et al., 2013), (Gauttam et al., 2013), (Wang et al., 2021), (Chen et al., 2018), (Dutta et al., 2021), (Lombardi et al., 2020), (Viana et al., 2017), (Chethana et al., 2016), (Shobha et al., 2015), (Mondal et al., 2020; Sasirekha et al., 2012), (Aradhya et al., 2021; Lyu et al., 2018), (Qi et al., 2011), (Huang et al., 2019), (Bezmaternykh et al., 2020), (Neha et al., 2012; Srivastva et al., 2013; Konya et al., 2010), (Shivakumar et al., 2005), (Jo et al., 2020), (Guru et al., 2015; Zhao et al., 2019), (Viana et al., 2017; Fourure et al., 2017; Chethana et al., 2016; Shobha et al., 2015; Mondal et al., 2020; Sasirekha et al., 2012; Aradhya et al., 2021; Lyu et al., 2018), (P. Agrawal et al., 2012), (R. Patel et al., 2015), (A. Kumar et al., 2014).]

The authors have proposed (B.S. Mamatha et al., 2014) a method to extract text from images with complex background which can be achieved using an edge-based text extraction algorithm based on the fact that edges are reliable feature of text regardless of font sizes, styles, color/intensity, layout, orientation, etc. Experimental results show that this method is very effective and efficient in localizing and extracting text-based features.

Using the proposed method, separating texts from a textured background with similar color to texts is performed (S. Minaee et al., 2017). Experimentation is carried out with their own dataset containing 300 image blocks in which several challenges like manually generated images by adding text on top of relatively complicated background. The proposed algorithm is robust to the initialized value of variables.

In this survey chapter, different issues like text detection, segmentation, and recognition natural scene images are discussed (U.B. Karanje et al., 2014). Comparison of different text detection methods based on the Maximally Stable Extremal Regions (MSER) is highlighted followed by advantages and disadvantages. From the survey, it is observed that detecting and recognizing text from natural scene images is more difficult task than all other existing methods. Even though there are many algorithms, no single unified approach can fit for all applications.

For text extraction in complex natural scene images, different methods based on color and gray information are proposed (H.-R. Byun et al., 2002). The proposed method works even if the document containing skew and perspective of candidate text regions. The method is tested in 128 natural scene images that are captured in various places such as in schools, hospitals, subway stations, and streets. The dataset is classified into two categories, simple and complex from the experimentation, color-based method gives better results than gray-based method for complex images but it has more false detections. The gray-based method has better performance for simple images. The combination of both the methods gives better results than that of each method [(Z. Ibrahim etal., 2008), (L. Sun et al., 2015), (Y. Song et al., 2017)].

More number of works on classification of text and non-text information can also be studied [(Chaithanya et al., 2019), (Tran et al., 2015),(Arvind et al., 2006), (Ghosh et al., 2018), (Puri et al., 2016), (He et al., 2019), (Lee et al., 2018), (Mishra et al., 2018), (Chen et al., 2007), (Diligenti et al., 2003), (Liu et al., 2021), (Hu et al., 1999). (Shirdhonkar 2010), (Zagoris et al., 2014). (Augusto et al., 2017), (Bhowmik et al., 2018), (Le et al., 2016), (Dhandra et al., 2010), (Saxena et al., 2019), (Kumar et al., 2016), (Thangaraj et al., 2018), (Blessieet al., 2019), (Kowsari et al., 2019), (Lin et al., 2017), (Kasar et al., 2013), (Gupta et al., 2019), (Banerjee et al., 2012). (Bhavani et al., 2021), (Gatos et al., 2005), (Ibrahim et al., 2008), (Gilani et al., 2017), (Bavdekar et al., 2015), (Riba et al., 2019), (Liu et al., 2013), (Li et al., 2018), (Ghosh et al., 2022), (Julca-Aguilar et al., 2017), (Ikonomakis et al., 2005), (Saha et al., 2019), (Schreiber et al., 2017), (Zhao et al., 2019), (Chen et al., 2011), (Tupaj et al., 1996), (Hao et al., 2016), (Kavasidis et al., 2018), (Schreiber et al., 2017), (Chen et al., 2007), (Kavasidis et al., 2018), (Zeiler et al., 2014), (Simonyan et al., 2014), (Gilani et al. 2017), (Okun et al., 1999), (Moll et al., 2008), (Nayef et al., 2015), (Fletcher et al., 1988), (Tombre et al., 2002), (Kavasidis et al., 2018), (Yi et al., 2017), (Ren et al., 2015), (He et al., 2015), (Everingham et al., 2010), (Lin et al., 2014), (Deng et al., 2009), (Girshick et al., 2015), (He et al., 2016), (S. Ghosh et al., 2018), (F.D. Julca-Aguilar et al., 2017), (Mandivarapu, et al., 2021)]

It is also important to note that no attempt to extract information from printed or handwritten bilingual document images. In India, most office documents, Advertisement boards, Inauguration Boards, Direction Boards, Answer Scripts (which are half printed and partially handwritten) are bilingual, with various text and non-text separation of document pictures. This presents new hurdles in the field of document image analysis, which has motivated us to continue this research work.

1.6 Challenges

From the literature review, we find that many challenging issues are still exists in text extraction from document images: they are

1. Extraction of text from blurred images.
2. Text extraction from handwritten document images.
3. Text extraction from multi scripts.
4. Multi oriented text detection.
5. Text extraction from logos.
6. Text recognition and many more.
7. Text extraction from partially visible documents.

1.7 Motivation and Objectives

From the literature review, we find that many challenging issues are still existing in text extraction from images: they are

Text and non-text separation is an important processing step in any document analysis system. It then divides off-line printed/handwritten document images into several types based on the nature of the problems each finds, in an attempt to provide understanding of the various techniques presented in the literature.

The characterization of complex text and non-text document images, necessary for both image segmentation and text and non-text classification remains a difficult and challenging problem. Feature selection is a challenging process in designing image classification systems. The complexity in this is user should be able to achieve insight into how observations behave in the feature space, since this may lead to design of better feature extraction methods.

To the best of our knowledge, no work has been documented on extracting text and non-text extraction from printed/handwritten bilingual document images with several challenges. As a result, in our research, we propose developing a novel contour approach and applying bounding boxes to various sections of document images. The developed model can be employed with or without a priori knowledge of document images for both handwritten and printed documents, and it can extract various text and non-text document images if they exist. The proposed models efficacy will be tested on a large number of document images with various parameters. The following are the major objectives of the proposed research work:

1. To propose an efficient classifier for classification of text and non-text information.
2. To Develop an effective segmentation algorithm for extracting the text.
3. To design an algorithm for estimation and correction of skew angle.

1.8 Organization of the Thesis

In Chapter 2, details on different types of datasets on printed/handwritten document images which are publicly and also all the datasets which are captured by the mobile camera and scanned document images. The interest of this chapter is to make the readers familiar with datasets on bilingual printed/handwritten document images available in literature on various qualities.

In chapter 3, Initially, we take gray scale real-time office document images, and interpolation is used to improve an image's visual appearance, i.e., its quality. The visual appearance of such an image is obtained by resizing it with the bilinear interpolation method. When the interpolation is concluded, we use the fuzzy logic approach to improve it. In fuzzy logic, an image is partitioned, and each partition is considered a fuzzy window. The fuzzy window is enhanced by using mean and variance. Similarly, all fuzzy windows are enhanced, and finally, all fuzzy windows are summed. Fuzzification, inference engines, and defuzzification are the three main parts of fuzzy logic. In this way skew angle of all the enhanced documents are calculated. The results of extensive experimentation conducted are tabulated. The performance is evaluated using quantitative measures like Michelon Contrast (MC), Entropy, Peak Signal to Noise Ratio (PSNR), Structure Similarity Index Measurement (SSIM), Absolute Mean Brightness Error (AMBE), Mean Squared Error (MSE) and Normalized Root Mean Squared Error (NRMSE) as a parameter.

Chapter 1, gives introduction of Text and Non-Text printed/handwritten bilingual document image enhancement, segmentation and Classification. The state of art existed methods and brief survey methods are presented.

In Chapter 2, we just made an attempt to brief our effort towards creation of the datasets. The first dataset consists of printed/handwritten bilingual document images with multiple skews and logos, pictures, tables, equations, numbers, seal impressions.

In chapter 3, we have proposed an efficient approach for enhancement of real time document images. The proposed approach Fuzzy Logic(FL) approach perform better than the existing methods.

In chapter 4, we have presented an approach for signature extraction from a bilingual document images. The proposed approach is based on contour and blob's method. The proposed algorithm is tested a two different cases i.e., before enhancement and after enhancement. Logo extraction is done by using masking and median filter techniques. To measure the performance different performance metrics are used like, Accuracy, Precision, Recall and F1-score.

In chapter 5, we proposed a hybrid U-SegNet model which integrates both U-net and SegNet architectures. The performance is evaluated using metrics like accuracy, precision, recall and F1-score and comparison analysis is also conducted with other segmentation methods such as watershed method, Fuzzy C- means and U-net method.

In chapter 6, We have proposed U-net and component-based region network is a different method for analyzing features of document images, such as regions, bounding boxes, convex hulls, filters, and enhancements, when compared with existing methods. The performance metrics used to measure performance are Accuracy, Precision, Recall, and F1-Score. The proposed method is compared with existing methods. Our method performs well.

In chapter 7, the overall contributions of the thesis are mentioned along with the summary of the work followed by directions for future research work.

The superiorities of the newly proposed methods in terms of effectiveness and robustness are established in the respective chapters theoretically, experimentally and also by an extensive comparative analysis with the other well-known methods.

Chapter 2

Creation of Real-Time Datasets

2.1 Preamble

Data availability has increased significantly as digital technology has progressed in recent years. A dataset is a collection or set of data. This information is usually presented in a tabular format. Each column represents a different variable. We give visual definitions of datasets and their potential elements. We collect a different types of datasets to describe the structure and concepts in a dataset, as well as the relationships between them. We apply to the several existing datasets derived from document images.

According to research, the quality of datasets is essential to the objectives of empirical studies in printed and handwritten document image analysis. For many years, researchers have used publicly accessible dataset repositories in their research. Although datasets are important in research, few studies consider the quality of their datasets which may result in questionable results if the datasets have quality issues. Some types of data were observed to get quality issues, according to studies. These studies take into account quality issues such as noise, missing data, and incorrect data. However, one issue that has received little attention whether or not the data is ambiguous or incapable of being interpreted correctly (D. Baviskar et al., 2021).

The ultimate goal of data analysis is to evaluate the quality of datasets collected in printed and handwritten document images, office documents, newspaper, magazines, road side name boards, and street boards. As a first step, we need all of the documents scanned and captured using an HP laser multi-jet scanner with a resolution of 300 dpi, as well as in terms of cameras, there is a 16-mega pixel mobile camera with a high resolution 16-mega pixel primary camera with an f/2.2 aperture and a 2-mega pixel camera with an f/2.4 aperture. Phase detection auto focus is available on the rear camera setup. It has a 16-mega pixel front-facing camera with an f/2.0 aperture. It's a good way to figure out what datasets are meant to represent. We start by looking at existing datasets to see how they're organized. Researchers then generate precise descriptions for datasets and their potential characteristics. As previously said, we develop a conceptual model that specifies a datasets structure and concepts, as well as the relationships between them. Researchers will be able to decide whether a dataset contains enough information to be relevant for analysis using the Definitions and model.

In this chapter, we present a new dataset of sentiment classification printed and handwritten document images for bilingual literature documents. This dataset tends to vary from previous semantic similarity datasets in that it contains documents as well as examples of difficult sentiment classification problems found in the literature. This dataset enables researchers to investigate the characteristics of long-distance within-document images as well as cross-domain performance for the task of being generally motivated. Initially, the benchmark dataset was used for shared tasks on modeling unrestricted frequently deal. Almost all modern systems are entirely dependent on this data. However, domains are relatively narrow, focusing primarily on document images (office documents, street signs, and signboards) While datasets exist for other domains such as document images, scientific articles, and school exams, Kannada and English texts is one area in which robust data is lacking.

In comparison to other languages like Kannada and English, real-time document images in DIA is still in its early stages. We created a page-level printed/handwritten document image dataset (PPHD) of official standard documents, such as Office Documents, Signboards, Street boards, Scanned Booklets, and Name boards, in this work. PPHD is made up of tens thousands of text and non-text document images from various parts of Karnataka. We also present the printed/handwritten benchmark results document identification (PHDI). Apart from document image detection, the dataset can be utilized in various of other document image analysis applications, including text-line segmentation, word segmentation/recognition, word spotting, handwritten and machine printed text separation, and writer identification.

2.2 Dataset of Bilingual Documents

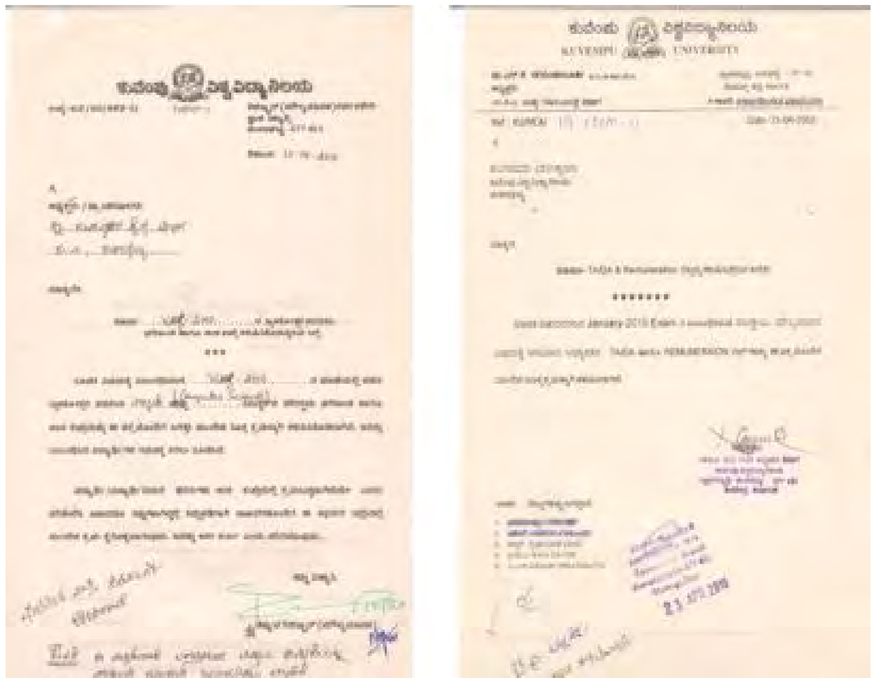
In this section, we present details on real time of a dataset of bilingual document images containing two languages i.e., one is the state official language Kannada, second one is national language and the other one is English which is a common official language at global level. in this case we have categorized the text and non-text separation of document images dataset into two types:first one is synthetically collected different places in common people, where we asked the individual to collected from outside in two different varieties of text and non-text documents i.e., Kannada and English with different orientations. In this way, a of total of 10000 document images were captured from different places. A sample of this synthetically created bilingual document images with multiple distortions, blurred document images, tilt letters, different image sizes and multiple skews.

Training data is collected from a variety of locations, such as streets, government offices, road or national highway side recorded photographs, private or government sector campuses dataset, and test data is collected from text and non-text images. All of the consistent standard images were set aside as a test dataset to ensure that the methodologies developed with this dataset generalize well to new educational images and possibly other fields. We also offer a baseline system based on a standard deep neural architecture and explore how to deal with the challenge of limited training data. Large datasets of labeled images, such like Image Net, are propelling convolution neural networks (CNNs) further in machine learning. Many information retrieval visuals, such as charts, tables, and diagrams, are made using software rather than photography or scanning.

A robust classifier of official document images illustrations offers wide range of applications in information extraction. Natural language systems could choose images by kind depending on what information a user is looking for, and techniques could immediately reveal filters by expected label. Additional analytic systems might be employed to extract more information from an image so that it can be indexed by classes. Software is used to design the majority of standard official images, which are qualitatively different from photos and scans (Sowmya Vajjala et al., 2020).

2.2.1 Five different datasets:

Dataset-I Office document images



(a) (b)

Figure 2.1: Sample (a) Sample (b)

Dataset-II Street board document images

(a) (b).

Figure 2.2: Sample (a) Sample (b)

Dataset-III Inauguration board images

(a) (b)

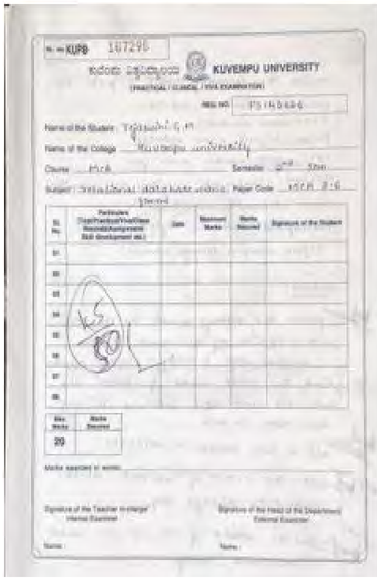
Figure 2.3: Sample (a) Sample (b).

Dataset-IV Sign board images

(a) (b)

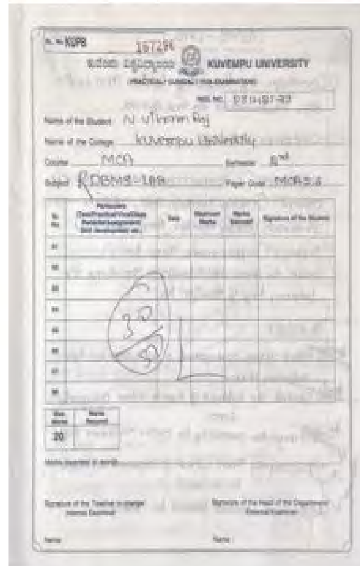
Figure 2.4: Sample (a) Sample (b).

Dataset-V Answer scripts



KUPB 167296
KUVEMPU UNIVERSITY

(a)



KUPB 167296
KUVEMPU UNIVERSITY

(b)

Figure 2.5: Sample (a). Sample (b).

Table 2.1: Different types of Datasets.

Data sets	Class 1	Class 2	Class 3	Class 4	Class 5	Total
1	Office Documents	Advertisement Boards	Inauguration Boards	Direction Boards	Answer Scripts	Over-all datasets
Total	2000	2000	2000	2000	2000	10000

2.3 Datasets

The ideal data set would allow researchers to project the performance achieved experimentally to the application domain represented by dataset.

Datasets allow researchers to train and test algorithm on significant numbers of data items and to compare performance on specific images. These are two main areas of

OCR research (off-line documents, on-line documents) which can be either handwritten or machine written.

Office documents typically contain text and non-text, text may contain alphanumeric characters, where as non-text contains logo, pictures, etc.,

An important condition in the design of datasets in this area is how well the subjects are gathered. subject should be chosen from the same population and date gathered under same condition.

The data sets gathered for off-line document processing are explained below in detailed.

1. Office documents: are the records kept to show details about sales and purchase or organization makes. They include invoice, credit note, debit note, receipt , delivery note, catalog, user guide, spread sheets bills financial statements etc.,

2. Academic documents: which includes marks sheets, transcripts, thesis, paper charts, journals, manuscripts etc.,

3. Boards: These are various kinds of board such as advertising boards, store boards, which again contain information in different languages and may contain logos which is non-text information.

4. Sign boards: on street boards, sign boards contain different information in multilingual form. The analysis of boards and processing into digital forms helps the user to understand the board easily which might be in different language. User can translates it into another known language to understand its meaning.

5. Answer scripts: which contain student registration number, which can be combination both numbers and alphabets it also contains questions and Answer written

by students in one or two languages. Analysis and processing of answer scripts helps in digital evaluation of answer scripts which further reduce the load of human intervention.

2.4 Conclusion

In this chapter, we just made an attempt to brief our effort towards creation of the datasets. The first dataset consists of printed/handwritten bilingual document images with multiple skews and logos, pictures, tables, equations, numbers, seal impressions. In this case. We have five different types of dataset classes Table 2.1; each classes consisting of 2000 document images with forwarded notices. The second dataset consists of 2000 bilingual printed/handwritten document images of different office documents. The third dataset consists of 2000 street boards images. Fourth and fifth dataset consists of each classes have 2000 sign boards and scanned booklets document images. Further, we also created a text and non-text document image datasets by extracting words, logos, tables, signatures from the all classes datasets. Here we extracted different bilingual printed and handwritten document images totaling a images dataset of 10000 document images.

Chapter 3

Enhancement of Document Images Using Machine Learning Approach

3.1 Preamble

The goal of image enhancement is to make an image more effective for a certain jobs, such as making a more individually pleasant image for human sight. The quality of an image as perceived by a person can be improved using image enhancement techniques. Because many real time document images on a color display provide insufficient information for image interpretation, these techniques are quite beneficial. There is no conscious effort to improve the image's integrity in comparison to some ideal form. Image quality can be improved using a variety of approaches. The most widely utilized techniques are contrast, stretch, density slicing, edge enhancement, and spatial and frequency domain. After correcting for geometric and radiometric distortion, image improvement is tried. Image enhancement procedures are applied separately to each band of a multispectral image. Because of the precision and variety of digital processes, digital techniques have been proven to be more satisfying than photographic procedures for image enhancement.

Image enhancement techniques are frequently ad hoc, with little or no attempt to

Some parts of the materials in this chapter have appeared in the following research paper.

1. Ravikumar M, Shivakumar G and Shivaprasad B J. 2022. "Enhancement of Real Time Document Images Using Fuzzy Logic and Machine Learning Approach", Journal of Jilin University (Engineering and Technology Edition Vol:41 Issue:09:2022. (Scopus Indexed).

anticipate the real image deterioration process. This procedure has no effect on the data fundamental information content. Gray level and contrast adjustment, noise reduction, edge sharpening, filtering, interpolation and magnification, and pseudo coloring are all included. Frequency domain and spatial domain methods are the two types of image enhancing techniques. The former transforms the image into a two-dimensional signal and the enhances using the image's two-dimensional Fourier transform. The low-pass filter approach removes noise from the image, and high-pass filtering enhances the edge, which is a type of high-frequency signal, and clarifies the fuzzy image. The local mean filtering–based approach and the median filtering (take intermediate pixel value of the local neighborhood)–based methods are two common spatial domains–based algorithms that can be used to eliminate or weaken noise.

A document image is first preprocessed with decolorizing, denoising, filtering, or gray boosting, among other things, in a standard OCR system. The gray scale image is then converted to a binary image using a threshold value for ease of usage in the following phase. Characters in the document picture are then separated and normalized during the character segmentation procedure. Furthermore, specific character feature statistics can be extracted and utilized in the final recognition stage, which results in the production of ultimate character strings as the text content in that image. Marginal noise is found and removed using the suggested strategy, which involves three iterations of block identification using the Hu moments method and converting the neighbour pixel to the background pixel. Marginal noise is commonly found towards the edges of document images, resulting in a non-uniform lighting gradient. In geosciences, astronomy, facial reconstruction, multiple-description coding, resolution enhancement, and geographic information systems, interpolation is a common approach for image scaling.

The spatial domain refers to the image plane itself and methods in spatial domain are based on directly modifying the value of the pixels. Document image enhancement is required in many situations when analyzing the quality of documents like handwritten/printed text documents, answer booklets, Street Boards and inauguration board images that have become noisy and low-contrast after scanning. One of the most important issues in document image analysis is contrast enhancement. Various types of need-based analysis tasks become more difficult as a result of high or low contrast on images. The solution for such images is to improve them by reducing the noise and increasing text contrast. This is possible by incorporating point operations

Here we have compared three different methods for image enhancement for document Images (i) Contrast stretching and (ii) Histogram Equalization. A number of contrast measures were proposed for complex images as document images. During image acquisition the images are affected due to poor illumination, lack of dynamic range in the imaging sensor, or even wrong setting of a lens aperture etc., To overcome this, we have to increase the dynamic range of the gray levels in the image being processed.

In this work, we propose a method i.e. Fuzzy Logic approach to enhance the real time document images and the proposed method is compared with spatial domain and frequency domain methods. Based on the quantitative metrics result is measured and result shows the superiority of the proposed method.

This chapter is organized as follows: In section 3.2 proposed method is discussed. In section 3.3 we discuss the Results and Discussions also a result are given and finally conclusion is given in section 3.4.

3.2 Proposed Method

Based on a review of literature, we understand that, no work has been reported on fuzzy based approaches for printed/handwritten document images of bilingual document images. However, the literature also suggests that, different feature extraction methods and classifiers affect the performance of document images differently, with an intuition that fuzzy of different feature extraction methods as well as combination of classification algorithms may enhance real time document images identification accuracy, we make an empirical analysis for possibility of fuzzy based approaches for real time bilingual document images which work both at frequency domain and spatial domain.

In this section, we discuss the proposed method for enhancing real time document images and the block diagram of the proposed method is shown in the Figure3.1.

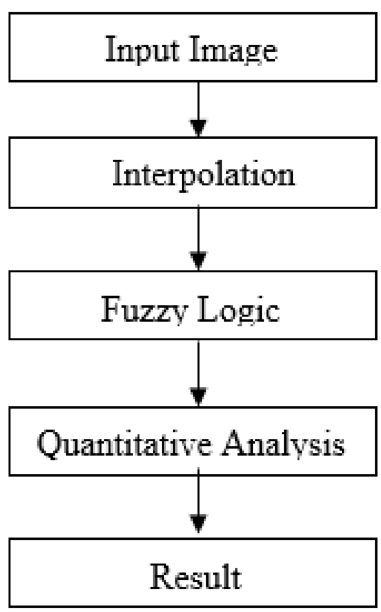


Figure 3.1: Block diagram of the proposed method.

Initially, we take grayscale real-time office document images, and interpolation is used to improve an image's visual appearance, i.e., its quality. The visual appearance of such an image is obtained by resizing it with the bilinear interpolation method (Puri Shalini et al., 2020). When the interpolation is concluded, we use the fuzzy logic approach to improve it. In fuzzy logic, an image is partitioned, and each partition is considered a fuzzy window. The fuzzy window is enhanced by using mean and variance. Similarly, all fuzzy windows are enhanced, and finally, all fuzzy windows are summed. Fuzzification, inference engines, and defuzzification are the three main parts of fuzzy logic (A. Thakur et al., 2015), which are given in the equation 3.1 and 3.2.

Here fuzzification is required to map the input image with fuzzy plane, vice versa for defuzzification 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 3.1.

$$\mathrm{Wij} = \frac{(P_{ij}(x,y))^{\gamma}}{\sum_{i=1}^{n}\sum_{j=1}^{m}(P_{ij}(x,y))^{\gamma}} \tag{3.1}$$

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 built as a sum of the transformed W_{ij} weights with degree of membership ψ_{ij}. The enhanced image is given by equation 3.2.

$$\Psi\mathrm{enh(f)} = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} X \Psi_{ij}(f) \tag{3.2}$$

where, ψ_{ij} (f) is image (f) before enhancement. ψ_{ij} (f) is image (f) after enhancement.

After the Fuzzy Logic process is completed, we compute quality of an images by using quantitative measure like Entropy, Peak Signal Noise Ratio (PSNR), Michelson Contrast (MC), Structure Similarity Index Measurement(SSIM) and Absolute Mean Brightness Error(AMBE), Mean Squared Error (MSE), Normalized Root Mean Squared Error (NRMSE) as parameter. Obtained results are tabulated in Table.[3.11 to 3.15] and corresponding the images are shown in the Figure [3.22 to 3.26].

In the next section, experimentation and results are discussed.

3.3 Results and Discussion

For the purpose of experimentation, we have used considered,real time scanned document images and captured images produced by the devices may have distortions such as blurred or noisy images. blurred document images, handwritten/printed text documents, answer booklets, Street Boards and inauguration board images, among other things. To enhance the real time document images, the different spatial domain enhancement methods are used like Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Gamma Correction(GC), Linear Transformation(LT), Log Transform(LogT) and Contrast Stretching(CS). Frequency domain methods focus on the image orthogonal transform instead of the image itself. Low Pass Filter(LPF), High Pass Filter (HPF), Gaussian Filter (GF), Non Local Means (NLM), Constrained Least Squares (CLS),Pseudo Inverse Filter(PIF) are used. After the experimentation, the results are shown in Spatial and Frequency domain the Figure 3.22 to 3.26.

To measure the performance of the images, we have considered quantitative parameters. They are Entropy,MC, PSNR, SSIM, AMBE,MSE and NRMSE. Table [3.1 to 3.10]. Various evaluation metrics were using NLM, LT and Fuzzy Logic(FL) processed image their respective spatial domain, frequency domain and fuzzy logic. The evaluation metrics used are described as follows. From table [3.1 to 3.5]., it is observed that the different real time dataset gives good result for Entropy, PSNR, AMBE and quantitative metric values are show in the Figures3.3, 3.5, 3.7, 3.9, 3.11, 3.13, 3.15, 3.17, 3.18 and 3.21, where the lower AMBE and higher Entropy, MC, SSIM, PSNR values indicates good quality image. Further, we need to extend our proposed method to improve MC and SSIM matrices.

Table 3.1 to 3.5 spatial domain and Table3.6 to 3.10, Frequency domain and Finally, Fuzzy Logic summarizes the various methods available in literature along with the proposed one. It is understood that unlike other, ours works on enhanced real time document images.

For experimentation purpose, we have considered our dataset into five different categories, they are

There are the five different datasets combined for one experimentation purpose and the total dataset containing five thousand images. Here we have applied these different approaches for enhancement and those images. They are spatial domain, frequency domain and fuzzy logic approaches. Here we have proposed fuzzy logic approaches which gives a good results and the proposed method compared with other two methods. Experimentation is discussed in these cases i.e., spatial domain and frequency domain methods.

$$[H] \text{ Entropy } = \sum_{i=1}^{C} -p_i * \log_2 (p_i) \tag{3.3}$$

$$[H] PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \tag{3.4}$$

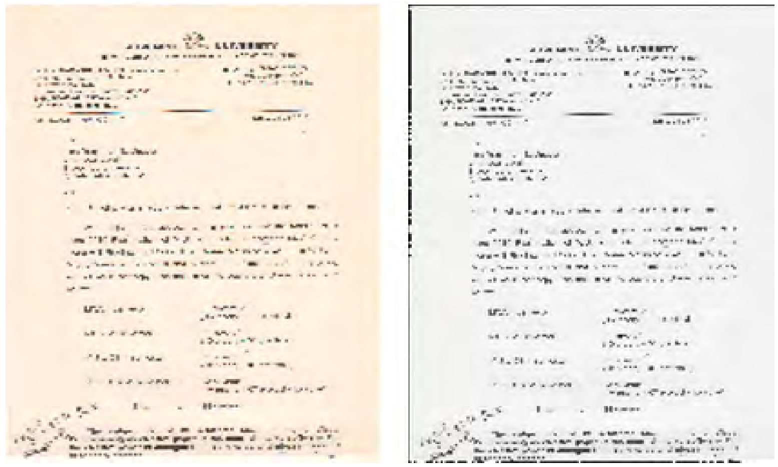
$$[H] \, \mathrm{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_1^2 + \mu_2^2 + c_1)(\sigma_1^2 + \sigma_2^2 + c_2)} \tag{3.5}$$

$$[H] MSE = \frac{1}{mn} \sum_{0}^{m-1} \sum_{0}^{n-1} \| f(i,j) - g(i,j) \|^2 \tag{3.6}$$

3.3.1 Case I. Spatial Domain Approach

In this section we have taken five real time datasets.

Dataset-I Office Documents:



(a) (b)

Figure 3.2: (a) Input Image. (b) Output Image.

Table 3.1: Comparison of Different quantitative methods.

Methods	Entropy	MC	PS-NR	SSIM	AM-BE	MSE	NRMSE
HE	16.88	1.0	0.7	0.0	233.27	55337.99	1.0
AHE	21.12	0.95	0.71	0.01	233.03	55225.44	1.0
Gamma	4.6	15.93	29.13	0.99	7.17	78.32	0.04
LogT	2.18	1.0	15.72	0.9	11.84	1744.02	0.18
CS	1.07	1.0	20.01	0.93	13.33	649.3	0.11
LT	**4.23**	**1.0**	**17.97**	**0.98**	**3.21**	**1038.6**	**0.14**

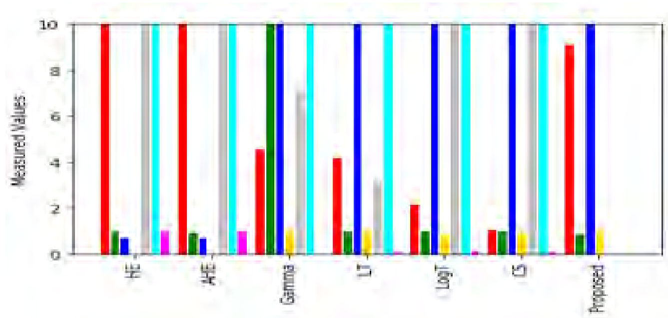


Figure 3.3: Graphical representation of different quantitative measures.

In this dataset, LT gives best result among all the remaining methods which is shown in Table 3.1.

Dataset II-Advertisement Boards

(a) (b)

Figure 3.4: (a)Input Image.(b)Output Image.

Table 3.2: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
HE	21.08	1.0	5.58	0.02	128.34	17997.54	1.0
AHE	21.77	1.0	5.58	0.02	128.3	17996.24	1.0
Gamma	6.94	1.0	18.93	0.94	28.31	831.39	0.21
LogT	5.61	1.0	8.96	0.8	88.8	8268.83	0.68
CS	6.86	1.0	13.49	0.69	48.14	2909.86	0.41
LT	**6.93**	**0.99**	**47.68**	**1.0**	**1.0**	**1.11**	**0.01**

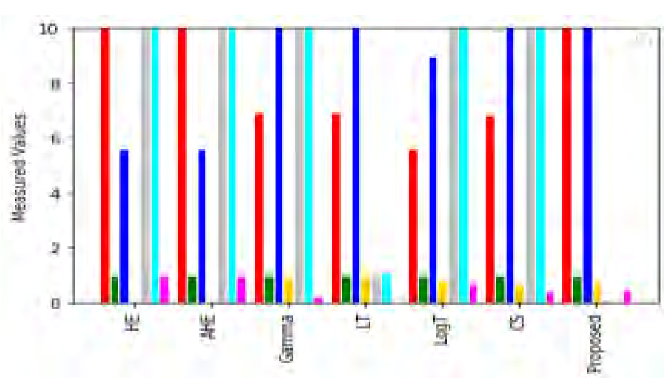


Figure 3.5: (a) Real time office document image enhanced document.

In this dataset, LT gives best result among all the remaining methods which is shown in Table 3.2.

Dataset-III Inauguration Boards

(a) (b)

Figure 3.6: (a) Input Image.(b) Output Image.

Table 3.3: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
HE	20.56	1.0	10.21	0.02	57.75	6196.11	0.99
AHE	21.78	1.0	10.2	0.02	57.87	6212.64	0.99
Gamma	5.92	1.0	21.86	0.77	19.89	423.31	0.26
LogT	6.51	49.8	6.71	0.47	115.23	13860.67	1.49
CS	2.4	1.0	16.72	0.24	32.02	1385.38	0.47
LT	**6.47**	**0**	**47.52**	**1.0**	**1.0**	**1.15**	**0.01**

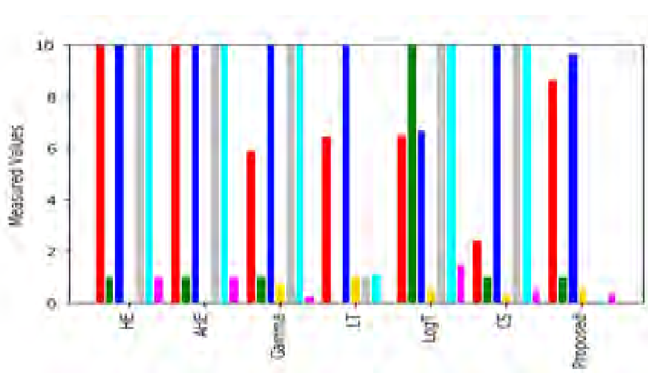


Figure 3.7: Graphical representation of different quantitative measures.

In this dataset, LT gives best result among all the remaining methods which is shown in Table 3.3.

Dataset-IV Direction Boards

(a) (b)

Figure 3.8: (a) Input Image.(b) Output Image.

Table 3.4: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
HE	21.29	1.0	4.33	0.0	144.69	2397.28	1.0
AHE	21.78	1.0	4.33	0.0	144.69	2397.28	1.0
Gamma	7.57	1.0	19.64	0.95	24.94	706.26	0.17
LogT	6.0	1.0	9.06	0.68	77.76	8070.99	0.58
CS	6.04	1.0	15.14	0.65	30.31	1992.29	0.28
LT	**7.47**	**1.0**	**22.49**	**0.97**	**1.0**	**366.53**	**0.12**

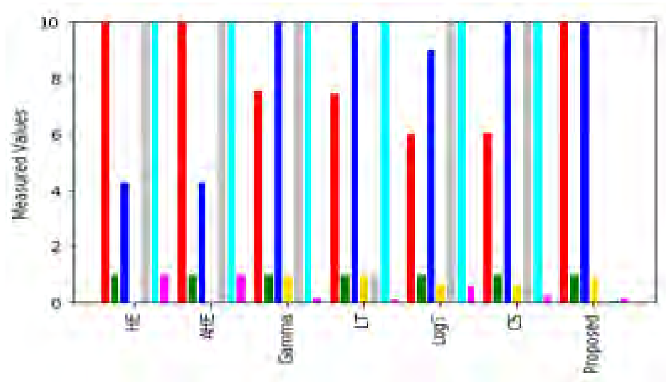


Figure 3.9: Graphical representation of different quantitative measures.

In this dataset, LT gives best result among all the remaining methods which is shown in Table 3.4.

Dataset-V Answer scripts

(a) (b)

Figure 3.10: (a) InputImage. (b) OutputImage.

Table 3.5: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
HE	15.4	1.0	0.46	0.0	240.26	58486.09	1.0
AHE	21.02	0.99	0.47	0.01	239.99	58361.84	1.0
Gamma	4.71	1.0	29.83	0.99	4.73	67.66	0.03
LogT	3.82	1.0	13.22	0.76	2.56	3097.1	0.23
CS	2.12	1.0	26.64	0.91	2.92	141.0	0.05
LT	4.87	1.0	13.95	0.87	12.21	2615.77	0.21

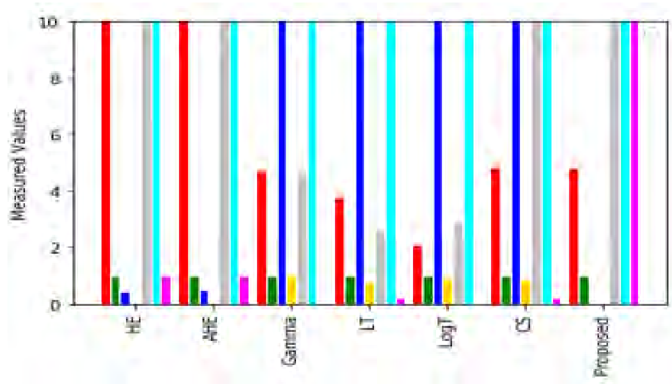


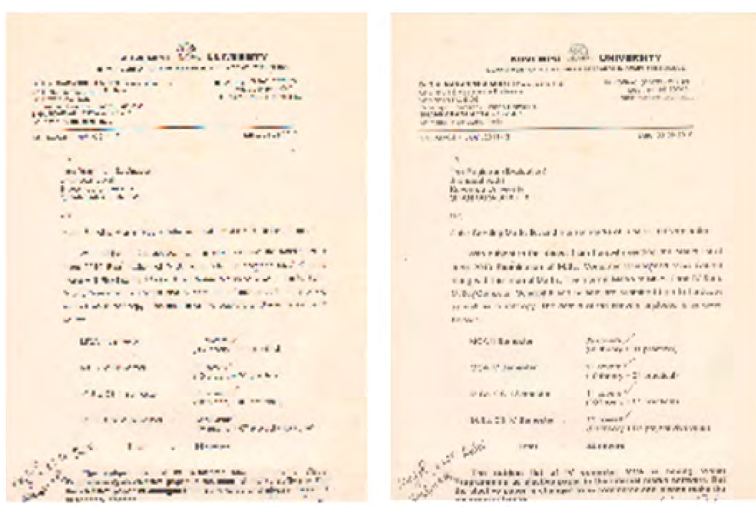
Figure 3.11: Graphical representation of different quantitative measures.

In this dataset, LT gives best result among all the remaining methods which is shown in Table 3.5.

3.3.2 Case II. Frquency Domain Approach

In this section we have taken five real time datasets.

Dataset-I Office Documents



(a) (b)

Figure 3.12: (a) Input Image.(b) Output Image.

Table 3.6: Comparison of Different quantitative methods.

Methods	**Entropy**	**MC**	**PSNR**	**SSIM**	**AMBE**	**MSE**	**NR-MSE**
LPF	5.41	4.4	26.76	0.94	0.6	137.12	0.05
HPF	7.62	1.0	4.16	−0.01	126.62	24969.61	0.68
GF	5.42	4.4	26.78	0.94	0.59	136.53	0.05
CLS	21.81	0.85	0.82	0.01	229.81	53874.7	1.0
PIF	21.81	0.85	0.82	0.01	229.81	53874.7	1.0
NLM	**5**.41	**1**.0	**20**.99	**0**.85	**0**.78	**5**17.33	**0**.1

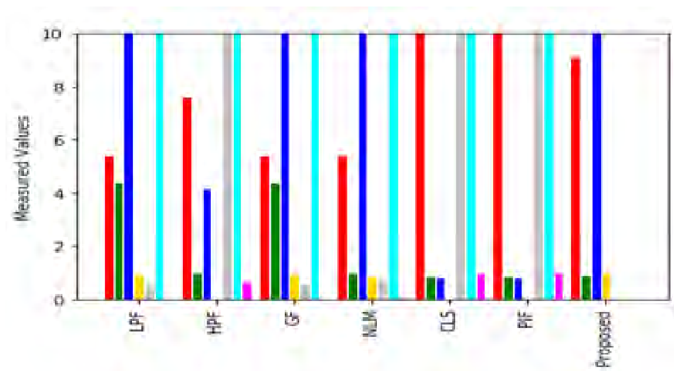


Figure 3.13: Graphical representation of different quantitative measures.

In this dataset, NLM gives best result among all the remaining methods which is shown in Table 3.6.

Dataset-II Advertisement Boards

(a) (b)

Figure 3.14: (a)Input Image. (b)Output Image.

Table 3.7: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
LPF	7.01	1.0	40.85	0.97	0.48	5.35	0.02
HPF	7.83	1.0	8.43	−0.02	5.01	9340.19	0.73
GF	7.01	1.0	40.87	0.97	0.49	5.33	0.22
CLS	21.81	1.0	5.72	0.02	126.05	17434.66	1.0
PIF	21.81	1.0	5.72	0.02	126.05	17434.66	1.0
NLM	**7.0**	**1.0**	**31.96**	**0.92**	**0.4**	**41.4**	**0.05**

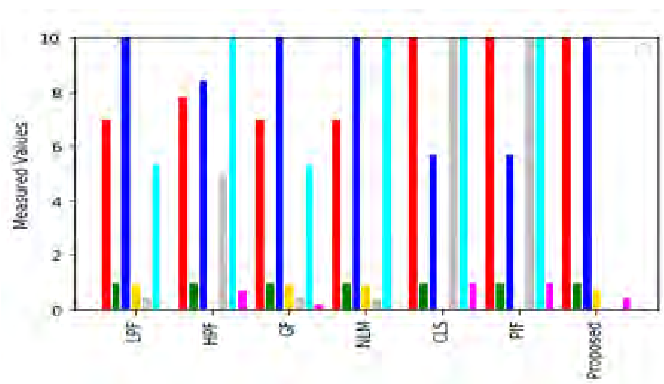


Figure 3.15: Graphical representation of different quantitative measures.

In this dataset, NLM gives best result among all the remaining methods which is shown in Table 3.7.

Dataset-III Inauguration Boards

(a) (b)

Figure 3.16: (a)Input Image. (b)Output Image.

Table 3.8: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
LPF	6.52	251.0	26.74	0.94	0.69	137.86	0.15
HPF	7.89	1.0	6.83	−0.01	62.38	1349.65	1.46
GF	6.53	251.0	26.75	0.94	0.69	137.34	0.15
CLS	21.81	1.0	10.18	0.01	58.19	6245.49	1.0
PIF	21.81	1.0	10.18	0.01	58.19	6245.49	1.0
NLM	**6.52**	**1.0**	**20.37**	**0.81**	**1.05**	**596.87**	**0.31**

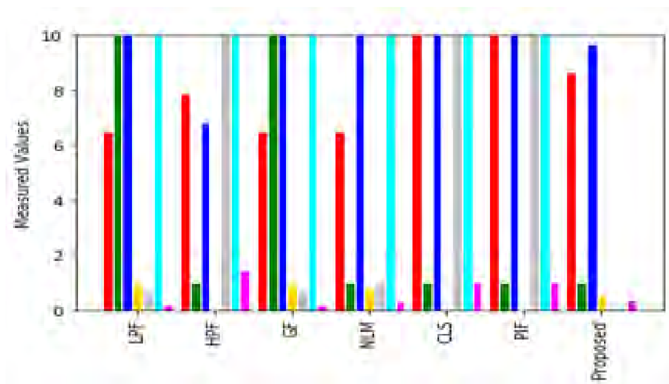


Figure 3.17: Graphical representation of different quantitative measures.

In this dataset, NLM gives best result among all the remaining methods which is shown in Table 3.8.

Dataset-IV Direction Boards

(a) (b)

Figure 3.18: (a)Input Image. (b)Output Image.

Table 3.9: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
LPF	7.67	126.0	26.25	0.88	0.51	154.11	0.08
HPF	7.98	1.0	7.69	−0.04	20.09	1107.41	0.67
GF	7.67	126.0	26.27	0.89	0.52	153.61	0.08
CLS	21.81	1.0	4.26	0.01	142.49	24372.7	1.0
PIF	21.81	1.0	4.26	0.01	142.49	24372.7	1.0
NLM	**7.67**	**1.0**	**19.34**	**0.68**	**0.44**	**756.78**	**0.18**

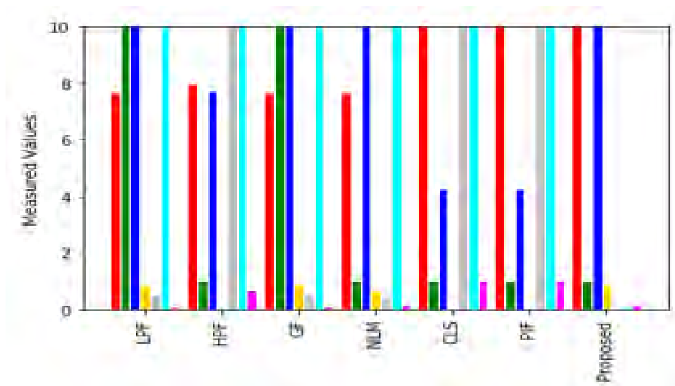


Figure 3.19: Graphical representation of different quantitative measures.

In this dataset, NLM gives best result among all the remaining methods which is shown in Table 3.9.

Dataset-V Answer Scripts

(a) (b)

Figure 3.20: (a)Input Image.(b)Output Image.

Table 3.10: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
LPF	4.63	1.0	31.79	0.96	0.42	43.05	0.03
HPF	6.37	1.0	2.51	−0.03	163.58	36508.69	0.79
GF	4.63	1.0	31.81	0.96	0.41	42.87	0.03
CLS	4.64	1.0	23.53	0.83	0.62	288.19	0.07
PIF	21.81	1.0	0.46	0.01	239.88	58484.44	1.0
NLM	**21.81**	**1.0**	**0.46**	**0.01**	**239.88**	**58484.44**	**1.0**

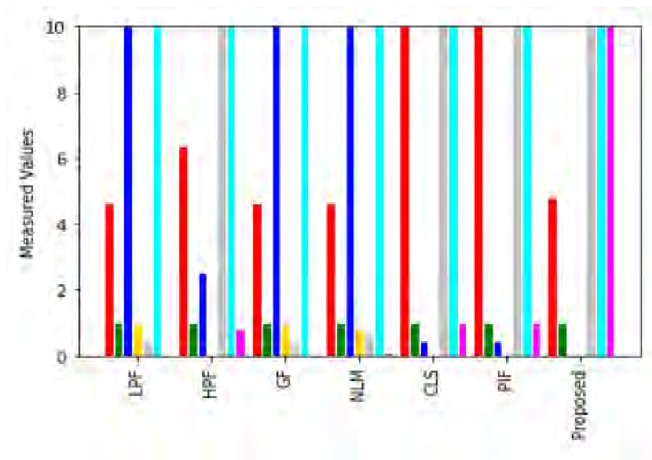
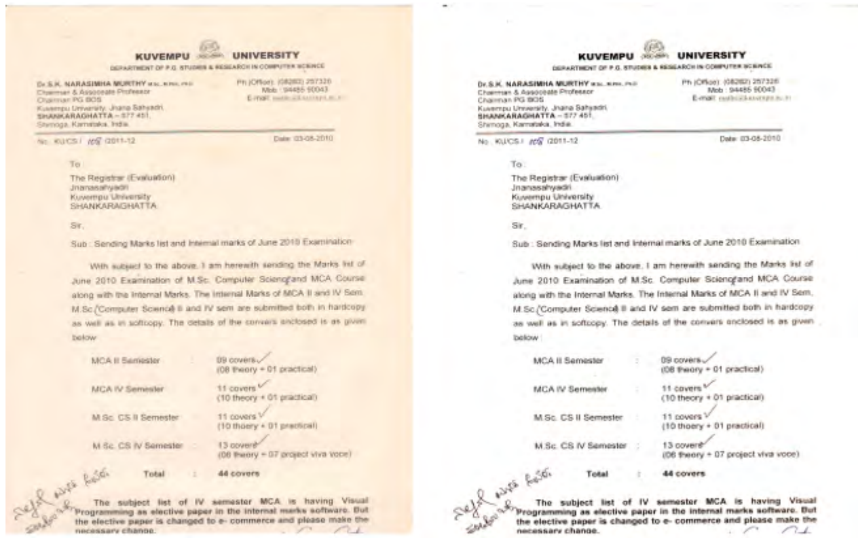


Figure 3.21: Graphical representation of different quantitative measures.

In this dataset, LT gives best result among all the remaining methods which is shown in Table 3.10.

3.3.3 Fuzzy Logic Method

Dataset-I Office Documents



KUVEMPU UNIVERSITY
DEPARTMENT OF P.G. STUDIES & RESEARCH IN COMPUTER SCIENCE

Dr.S.K. NARASIMHA MURTHY
Chairman & Associate Professor
Chairman PG BOS
Kuvempu University, Jnana Sahyadri,
SHANKARAGHATTA – 577 451,
Shimoga, Karnataka, India.

Ph (Office): (08282) 257326
Mob : 94485 90043
E-mail:

No.: KU/CS / /2011-12 Date: 03-08-2010

To:

The Registrar (Evaluation)
Jnanasahyadri
Kuvempu University
SHANKARAGHATTA

Sir,

Sub : Sending Marks list and Internal marks of June 2010 Examination

With subject to the above, I am herewith sending the Marks list of June 2010 Examination of M.Sc. Computer Science and MCA Course along with the Internal Marks. The Internal Marks of MCA II and IV Sem, M.Sc (Computer Science) II and IV sem are submitted both in hardcopy as well as in softcopy. The details of the convers enclosed is as given below :

MCA II Semester	:	09 covers (08 theory + 01 practical)
MCA IV Semester	:	11 covers (10 theory + 01 practical)
M.Sc. CS II Semester	:	11 covers (10 thoery + 01 practical)
M.Sc. CS IV Semester	:	13 covers (06 theory + 07 project viva voce)
Total	:	44 covers

The subject list of IV semester MCA is having Visual Programming as elective paper in the internal marks software. But the elective paper is changed to e- commerce and please make the necessary change.

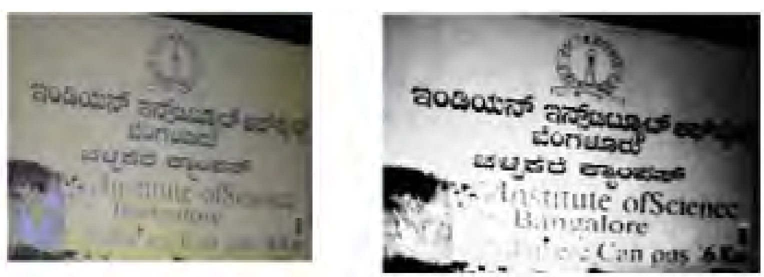
(a) (b)

Figure 3.22: (a)Input Image.(b)Output Image.

Table 3.11: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
FL	9.14	0.91	37.27	1.0	0.01	0.0	0.01

Dataset-II Advertisement Boards



(a) (b)

Figure 3.23: (a)Input Image.(b)Output Image.

Table 3.12: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
FL	13.84	1.0	12.93	0.78	0.08	0.05	0.43

Dataset-III Inauguration Boards

(a) (b)

Figure 3.24: (a)Input Image.(b)Output Image.

Table 3.13: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
FL	8.67	1.0	9.71	0.54	0.07	0.1	0.33

Dataset-IV Direciton Boards

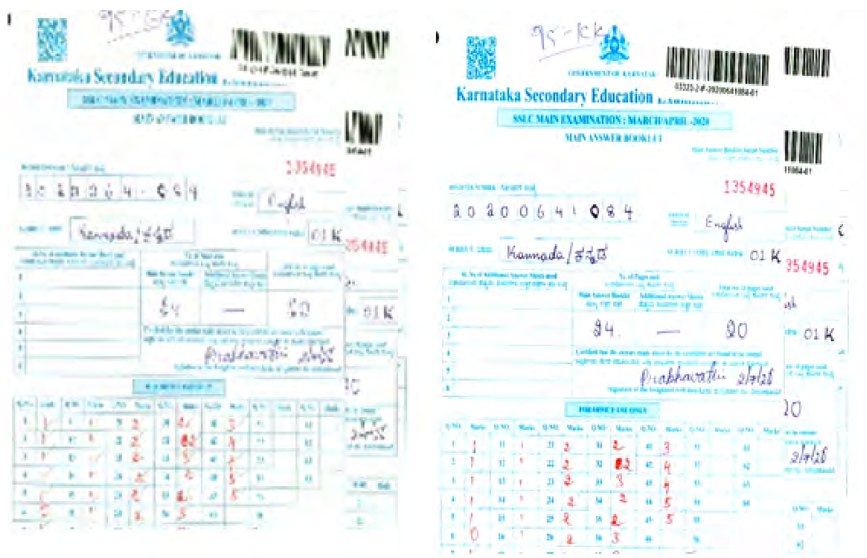
(a) (b)

Figure 3.25: (a)Input Image. (b)Output Image.

Table 3.14: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
FL	14.16	1.0	19.23	0.89	0.04	0.1	0.18

Dataset-V Answer Scripts



(a) (b)

Figure 3.26: (a)Input Image.(b)Output Image.

Table 3.15: Comparison of Different quantitative methods.

Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
FL	4.82	1.0	−47.39	0.0	232.14	54783.76	245.345

Table 3.16: Comparison of the results from proposed method for five different enhancement datasets.

Sl No.	Methods	Entropy	MC	PSNR	SSIM	AMBE	MSE	NR MSE
1	Balamurugan et.al.,[11]	4.23	1.0	17.97	0.98	3.21	7103.6	70.14
2	Antoni et.al.,[46]	5.41	1.0	20.99	0.85	0.78	517.3	0.1
3	Ravinder et.al.,[15]	16.88	1.0	0.7	0.0	233.27	5446.4	1.0
4	**Spatial Domain**	4.23 6.93 6.47 7.47 4.87	1.0 0.99 0 1.0 1.0	17.97 47.68 47.52 22.49 13.95	0.98 1.0 1.0 0.97 0.87	3.21 1.0 1.0 1.0 12.21	1038.6 1.11 1.15 366.53 2615.7	0.14 0.01 0.01 0.1 2 0.21
5	**Frequecy Domain**	5.41 7.0 6.52 7.67 21.81	1.0 1.0 1.0 1.0 1.0	20.99 31.96 20.37 19.34 0.46	0.85 0.92 0.81 0.68 0.01	0.78 0.4 1.05 0.44 239.8	517.33 41.4 596.87 756.78 5848.4	0.1 0.05 0.31 0.18 1.0
6	**Proposed (Fuzzy Logic)**	**9.14** **13.84** **8.67** **14.16** **4.82**	**0.91** **1.0** **1.0** **1.0** **1.0**	**37.27** **12.93** **19.71** **19.23** **−47.39**	**1.0** **0.78** **0.54** **0.89** **0.0**	**0.01** **0.08** **0.07** **0.04** **232.14**	**0.0** **0.05** **0.01** **0.01** **5478.7**	**0.01** **0.43** **0.33** **0.18** **245.3**

From Table 3.16, we have given the various approaches in enhancement of document images are given and sl.no.6, in the spatial domain approaches which contain five documents indicates that for experimentation we have used five different datasets shown in each row similarly frequency domain and the proposed method in Fuzzy Logic approach.

Table 3.17: Comparative analysis of Three different cases.

Cases		Entropy	MC	PSNR	SSIM	AMBE	MSE	NRMSE
Spatial Domain	D4	**7.47**	**1.0**	**22.49**	**0.97**	**1.0**	**366.53**	**0.12**
Frequency Domain	D3	**6.52**	**1.0**	**20.37**	**0.81**	**1.05**	**596.87**	**0.31**
Proposed (Fuzzy Logic)	D1	**9.14**	**0.91**	**37.27**	**1.0**	**0.01**	**0.0**	**0.01**

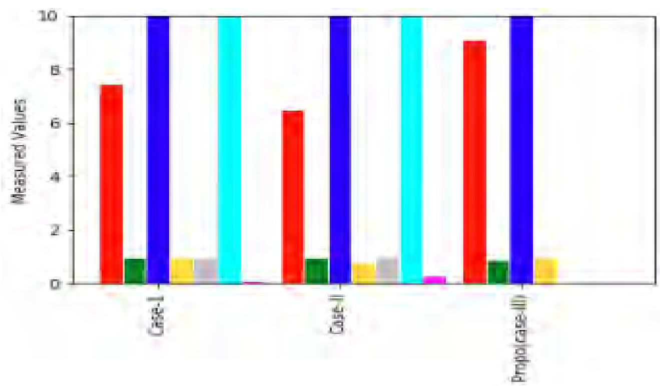


Figure 3.27: Graphical representation of different quantitative measures.

In this work we have used five different real time datasets for three cases and we have selected the best method from each case and case three i.e. fuzzy logic(FL) approach is the best method for enhancement of document images which is shown in the graph 3.27.

3.4 Conclusion

In this work, we have proposed an efficient approach for enhancement of real time document images. The proposed approach uses Fuzzy Logic(FL) method. Experimentation is carried out on our own five different datasets containing Five thousand complicated document images. The proposed method is compared with spatial and also frequency domain methods. Finally, Fuzzy Logic approach perform better than the existing methods.

Chapter 4

Extraction of Signature and Logo from Bilingual Document Images

4.1 Preamble

Section A

Signature is one of the most essential part generally studied for recognizing the biometric modalities. Nowadays signature detection and verification performs a dynamic part in the organizations somewhere security and confidentiality are the main anxieties. Especially in countries like India because these two features of the manuscript are the finest concepts for the retrieval of the contented information of the document. As the population is increasing day by day unique identification of a document can be done through signatures of the authors of an individual office. Signature extraction is the recent biometric proof of identity method, with high permissible recognition. Even uncertainty

Some parts of the materials in this chapter have appeared in the following research paper.

1. Shivakumar G., Ravikumar M., Shivaprasad B. J. and Guru D. S., 2022. “Signature Extraction from Bilingual Document Images Using Blobs Method”, In Modern Approaches in Machine Learning and Cognitive Science, pp. 283-294. (springer).
2. Shivakumar G., Ravikumar M., Shivaprasad B. J. and Guru D. S., 2022. “Extraction of Logo from Real Time Document Images Using Masking and Median Filter Approaches”, IEEE INCET, Technically Co-Sponsored by IEEE Bangalore Section and IEEE USA, pp. 01-07. (IEEE).

handwritten signature verification has been comprehensively considered in the preceding eras. The furthermost accurate techniques essentially always take development of dynamic features like acceleration, velocity, and the difference between up and down strokes.

Signatures may be responsible for amusing information about a person as they consist of exclusive belongings of human behavior, therefore they are used for detection/Identification determinations. The signature has been well-thought-out for bio-metric authentication in organizational documents, legal documents, etc. In a manuscript, the signature may be inspected by means of forensic document analysis experts for validating documents and to confine forged. It is a common organisational preparation currently to store and maintain large databases which is a determination to move towards a paperless office. Large quantities of administrative documents are often scanned and archived as images (e.g. Office documents dataset) without suitable directory information.

To segment documents from two layers: one layer supposed to contain printed text and other layer contain the handwritten parts. Such a segmentation problem has usual an inordinate deal of consideration in the literature for the reason that of the different processing methodologies for printed and handwritten texts. The objective is to apply corresponding techniques on the printed and handwritten parts. Many exploration works are going on for automatic online/offline signature verification and recognition (Guangyu Zhu et al., 2018; K S Radhika et al., 2014; Prakash H. N. et al., 2010; Prakash H N et al., 2009; Prakash H N et al., 2009).

Signature is a unique object and helps in indexing of large official papers stored in the database. Signature detection/verification from the document images and retrieving documents using the signature as a query is challenging task in the area of document image exploration and processing. Nevertheless, these methods take on that the signatures are inaccessible and do not touch/overlap with other text in the document.

In a machine-printed manuscript that comprises a signature, there might be some printed texts that may touch and/or overlap the signature. As a sign of that, such run-through has generated a remarkable demand for robust conducts to access and manipulate the information that these images contain. Obtaining information resources relevant to the query from such repositories is the main objective of document retrieval.

A sample scanned document from the Office document dataset So, signatures could be used as key data for searching and retrieval of documents. Thus, the handwritten signature will undoubtedly add an advantage for document indexing and searching. Since a signature is treated as non-text, it will be useful in the classification of text and non-text information from given document images.

4.2 Proposed model

In this section we discuss the proposed methodology. The illustration of the proposed method is shown in Figure 4.1.

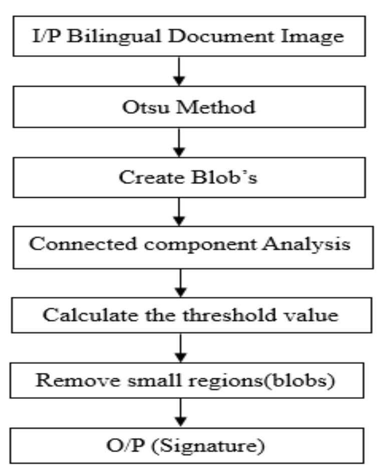


Figure 4.1: Block diagram of the proposed method.

At the beginning, bilingual document is considered for experimentation. Since the documents is real time, collected from different government offices it may contain noises, low resolution, may be blurred. Hence preprocessing is needed. To enhance such documents we use Otsu method. After this the document is free from noises. Normally signatures are irregular in shape containing different features like strokes, curves, edges, etc., to store all these features blob's were created. Since our intension is to extract signature from the document, it is necessary to separate foreground from back ground.

Since signatures generally have different parts, there may be some discontinuity in them. To fill up the discontinuity we are applying a method called connected component analysis. Normally area of the signature is more when compared with all the blobs and these small blobs are removed using region props, which is performed by fixing some threshold values, and finally the area i.e, a blob which is left in the documents is considered as a signature.

Since signature is having its own features when compared with text, by applying thresholding concepts background and foreground are separated. Here signature is considered foreground which is based on the threshold value and thus signature is considered as a blob. Some times there may be the same size of signature and non-signature information apart from this some small blobs also be created. Using connected component analysis any discontinuity present in the signature is filled. Normally area of the signature is more when compared with all the blobs and these small blobs are removed using region props. In order to eliminate smaller areas called blobs, region props are used which is performed by fixing some threshold values and finally the area i.e, a blob which is left in the documents is considered as a signature. The proposed algorithm extracts signatures from a bilingual document with any orientation in a real-time scenario even if most of the documents containing more than one signature at different locations.

4.3 Results and Disussion

For experimentation, we have created our own dataset containing 150 document images. The images are given to the proposed algorithm and performance is evaluated by measuring the Accuracy, Precision, Recall and F1-Score, Jaccard similarity, Dice and intersection over union as parameters. The proposed method gives good results compared to all the existing methods shown in Table 4.1.

Table 4.1: Different measuring parameters used for signature detection.

Methods	**Parameters**						
	Accuracy	**Precision**	**Recall**	**F1-score**	**Jaccard similarity**	**Dice**	**Intersection over union**
Histogram(Napa et al., 2013)	0.38	0.64	0.39	0.41	0.279	0.32	0.19
Contour(Aravinda et al., 2019)	0.43	0.59	0.44	0.42	0.250	0.47	0.30
Surf (Pal et al., 2012)	0.47	0.63	0.47	0.50	0.373	0.38	0.23
K-means(Alpana et al., 2018)	0.49	0.70	0.49	0.54	0.437	0.30	0.18
LBP (Tejas Jadhav et al., 2019)	0.62	0.73	0.62	0.66	0.571	0.40	0.25
Proposed	**0.84**	**0.94**	**0.88**	**0.91**	**0.829**	**0.48**	**0.32**

From the Table 4.1, it is observed that the proposed method gives good signature detection results for all metrics used. Results are shown in the Figure 4.2 to 4.11 and the corresponding graphical representation for all the techniques is plotted in graph which is shown in the Figure 4.12.

Figure 4.2: (a) Monolingual document images with three different signatures. (b) output.

Figure 4.3: (a) Bilingual document image with one signature. (b) Output.

(a) (b)

Figure 4.4: (a) Bilingual document image with single signature. (b) Output.

(a) (b)

Figure 4.5: (a) Monolingual handwritten document image with one signature. (b) output.

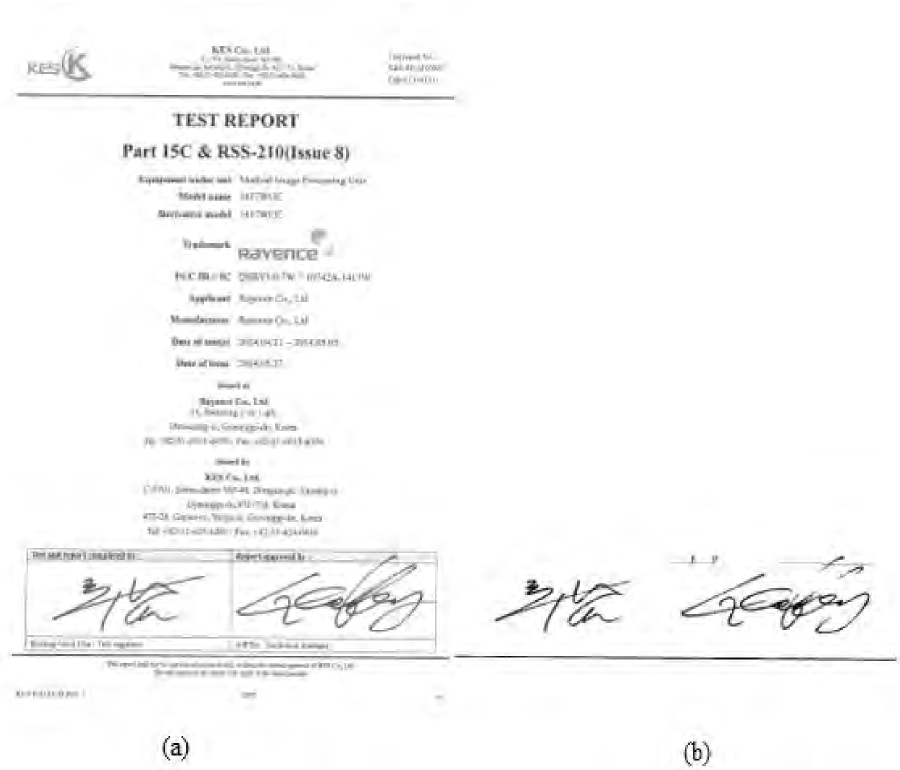


Figure 4.6: (a) Monolingual document image with two signatures. (b) Output image.

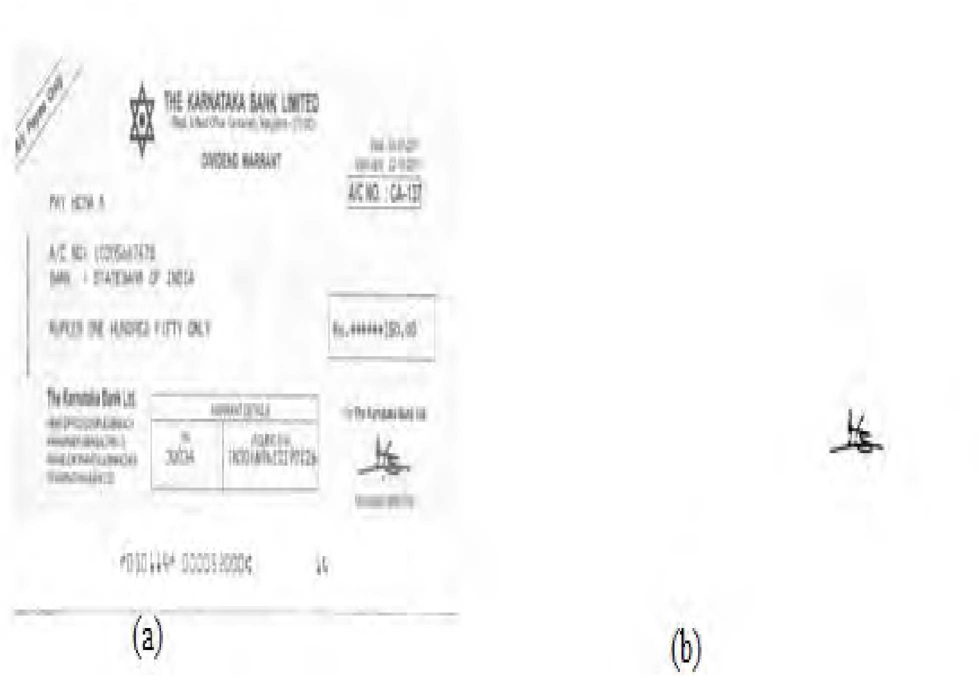


Figure 4.7: (a) Bilingual document with signature. (b) Output image.

Figure 4.8: (a) Document image with signatures. (b) Output image.

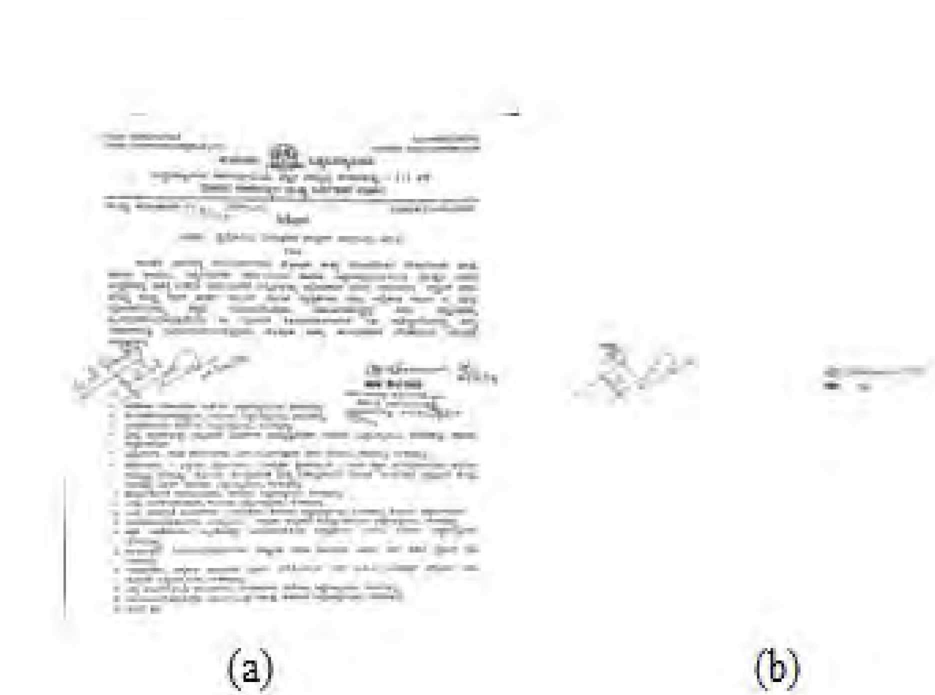


Figure 4.9: (a) Bilingual document with two signature. (b) Output image.

Figure 4.10: (a) Bilingual document with signature. (b) Output image.

Figure 4.11: (a) Document image with three signatures. (b) Output image.

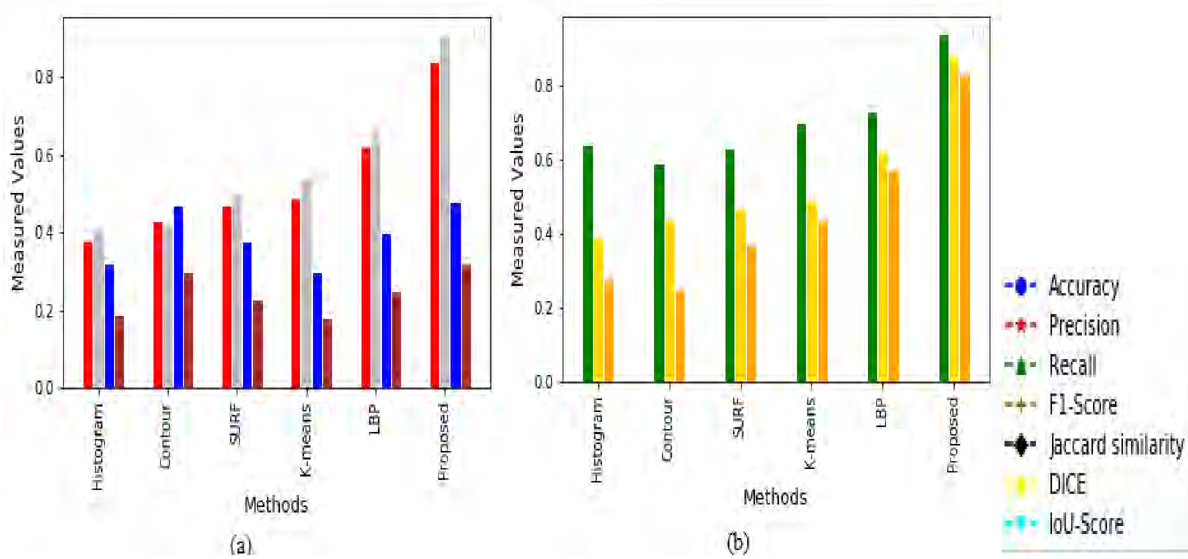


Figure 4.12: Graphical representation of signature detection for different techniques.

Section B

A logo is the visual representation of a company or organization, which forms the foundation of its identity. It is composed of name, symbol, monogram, emblem, and trademark, designed for easy and definitive recognition by public. Generally, logo is an essential part studied for recognizing the representation modalities. Recent study reveals an initiative taken to organize and verify the admired records using the computer logos in a particular location, which essentially contains collection of objects. A logo is a key component of a complex model for identification that must be applied fundamentally to form organization's communications. The creation of integration of logos into the proper technological scheme therefore is among the most difficult and important fields of graphic design. Research will continue to expand it to include access to sensitive documents such as scanned papers, documents by the university, governmental certificates, etc., large discoveries the corresponding institute. Every organization has a unique logo, but the size, color, texture, and pattern of it may vary (Vaijinath et al., 2017; Umesh et al., 2015).

Given an image of the document, identifying and differentiating that logo becomes a major task in retrieving and identifying documents based on logo. Logo identification helps primarily in the optimization of document images and improves the user's ease of authenticating documents contained in the database. Therefore, many researchers in the field of document image processing and document image analysis have drawn attention to the identification of a logo and its separation from the document. Furthermore, we deal with the problem of logo detection and include an ideal solution in this article (Umesh et al., 2016).

Nevertheless, the isolated or overlapping printed text present in the logo is difficult to extract. In a machine-printed manuscript that comprises a logo, there might be some printed texts that may touch or overlap the logo. As these difficulties are mentioned in the above case it is mandatory to access and manipulate the information present in the logo, obtaining data resources relevant to the query from such repositories is the main concept of document retrieval. A sample of a printed document images with a logo. So, the logo could be used as key data for searching and retrieval of documents. Since a logo is treated as non-text, it will be useful in the classification of text and non-text information from related document images. In Section 4.4, we present the proposed technique, in Section 4.5, we conducted Experimentation achieved by the proposed procedure, and finally in Section 4.6, the paper is concluded.

In the following we discuss the proposed method for logo extraction.

4.4 Proposed Methodology

In this part, we will discuss in detail about the proposed methodology. Figure 4.13, describes flowchart for logo extraction from document image.

In this section, we discuss the propose methodology and the block diagram of the proposed method is given Figure 4.13. Initially the input image is converted to binarization. Since we have collected document images and these image were captured through mobile camera as with different resolutions. Because of different resolution the quality of the images is also varies (some have good quality and some poor quality). Hence it is necessary to improve the quality of the image and can be performed using enhancement technique. In this work we have extracted logo from the document using two different approaches that is with enhancement and without enhancement. Once image is binarized and the binarization process can be performed using threshold level.

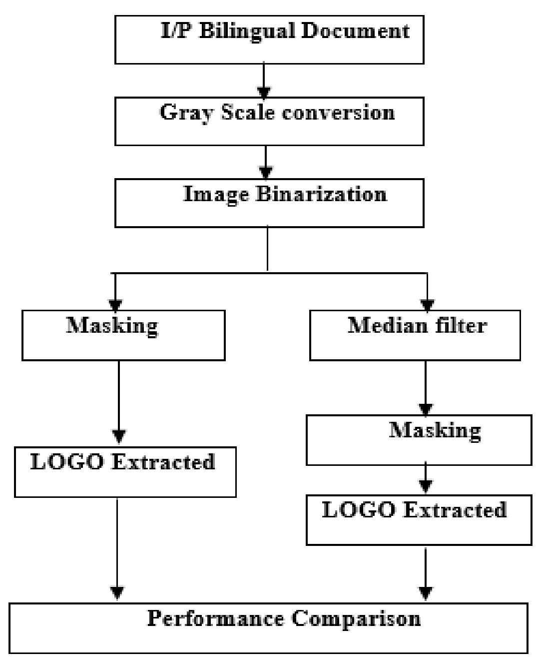


Figure 4.13: Flowchart of proposed method.

$$f(x,y) = \begin{cases} 0, & \text{if f(x,y)<T} \\ 1, & \text{if f(x,y)}\geq\text{T} \end{cases} \tag{4.1}$$

In this work our main intention is to extract logo from document image. The given image contain both text and non-text logo is a non-text and area of the text is normally smaller than are non-text in order to identify this larger area we are using a method called masking. Using masking the larger area from the document is extracted & masking is also used for segmentation purposes. In this way logo is extracted. In another approach after binarization we need to remove the noises for enhancement purpose. Here we have used median filter and removes impulse noise and it is denoted by

$$\hat{f}(x,y) = \frac{1}{mn} \sum_{(s,t)\varepsilon S_{xy}} g(s,t) \tag{4.2}$$

Noise removal: This step applies median filter to remove non-logo part in the document. Mainly it removes impulse noise from the image after masking operation. It keeps much of the logo part as it is in the document and converts remaining part of the document into background color.

Logo extraction: Now we draw a bounding box to the output obtained from the previous steps. This helps to find out location of a logo in the binarized image. The co-ordinate points of the bounding box in the barbarized image are used to extract a logo from original gray scale/color image.

Once the image is enhanced we go for masking as explained above and logo is extracted. After this we have compared both the approaches.

4.5 Experimentation

In this section, we will discuss the result obtained from the proposed methodology. Experimentation is carried out on our own data set of 500 document images containing both printed/handwritten which may be monolingual or bilingual. In this experimentation we have considered two different cases. Case one is for without enhancement and the other with enhancement.

4.5.1 Case-I: Logo Extraction Without Enhancement.

For experimentation, we have made our own dataset containing 500 document images. The input image given the proposed algorithm and performance is evaluated by measuring the Accuracy, Precision, Recall, F1-Score, Specificity, Dice Co-efficient, and Jaccard similarity as parameters. The proposed method gives good results compared to the Masking and Median filtering methods. The tabulated results are shown in Table 4.2, From that table, it is observed that logo extraction without enhancement gives less results of compared with with-enhancement documents contain noisy, blurred. Distortion logo present in the documents.

From Table 4.2, shows that the proposed method gives good logo extraction without enhancement results for all metrics used. Result are shown in the Figure 4.14 to 4.17 and the corresponding graphical representation for all the techniques is plotted in graph which is shown in the Figure4.18.

Table 4.2: Different measuring parameters used for logo extraction using (without enhancement).

Methods	**Accuracy**	**Recall**	**F1-score**	**Precision**	**Specificity**	**Dice Coefficient**	**Jaccard similarity**
Histogram	0.46	0.56	0.47	0.50	0.30	0.30	0.397
Contour	0.42	0.53	0.42	0.45	0.24	0.29	0.243
Surf	0.56	0.65	0.56	0.60	0.31	0.25	0.527
SVD	0.59	0.61	0.59	0.60	0.36	0.39	0.540
Proposed	**0.86**	**0.87**	**0.86**	**0.86**	**0.45**	**0.45**	**0.851**

For Logo extraction without enhancement, the performance of the proposed algorithm is evaluated by using different measuring parameters like accuracy, precision, recall, F1-score, specificity, Dice-coefficient and jaccard similarity. The values are tabulated in Table 4.2 and corresponding results are show in the Figure 4.18.

4.5.2 Case-II: Logo Extraction with Enhancement.

In this case we have extracted logo form the document after enhancement and the corresponding results are show.

For the purpose of experimentation, we considered 500 printed/handwritten document images collected from mobile camera which are real-time datasets. To improve the document images, dissimilar improvement approaches like Low Pass filter(LPF), High Pass Filter(HPF), Gabour Filter(GF), and Bilateral Filter(BLF) are used. After the experimentation, results are shown in Figure 4.19. to 4.22 The tabulated outcomes are revealed in Table 4.3.

By using dissimilar enhance logo document image with BLF value, image is improved and consuming filters. The consequences be present exposed in the above tabulated images. The PSNR and AMBE value is calculated in Table 4.4. When without enhancement document image compare to with enhancement value of PSNR is 0.0287 and AMBE is

0.0627, it gives proper outcome, result is stately founded on PSNR as a parameter.

This paper logo documentation RGB image as input from our own database created at that moment transform the RGB image to gray scale image then select the actual logo from image extract the logo from office documents by means of accuracy. This paper documentation dissimilar images of different institution with different sizes to the module and this module yield only extracted logo from huge images. And the logo size is not unique dimension with full accuracy. In this paper relates Masking method to extract and match the logo. Median filter method is used to extract the logo from document images. But previous methods enhanced than this method.

From the Table 4.4, it is observed that the proposed method gives good logo recognition with enhancement outcomes intended for all metrics used. Results are shown in the Figure 4.19 to 4.22 and the corresponding graphical representation for all the techniques is plotted in graph which is shown in the Figure 4.18 The proposed method gives good result for Accuracy, Precision, Recall, F1-Score, Specificity, Dice Co-efficient, and Jaccard similarity, we need to improve the further enhancement working methods.

The proposed method is compared with other existing methods, proposed method gives good result. Finally, from the experimentation it is observed that logo extraction with enhance gives good result of 90% recognition accuracy.

Figure 4.14: (a) Bilingual document with Logo image. (b). Logo detected.

Figure 4.15: (a) Bilingual document with university prospect logo image.(b) Logo detected.

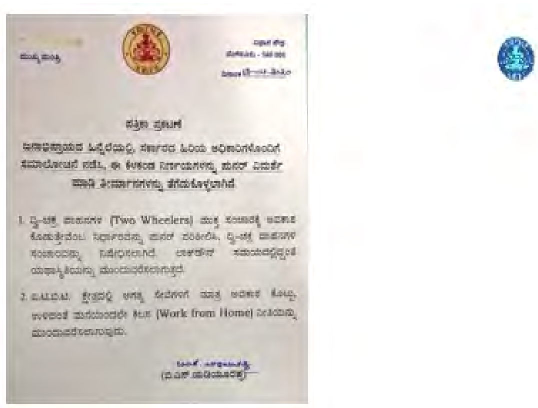


Figure 4.16: (a) Bilingual document with Karnataka Govt. office logo image.(b) Logo detected.

Figure 4.17: (a) Bilingual document with university logo image input image. (b) Logo detected.

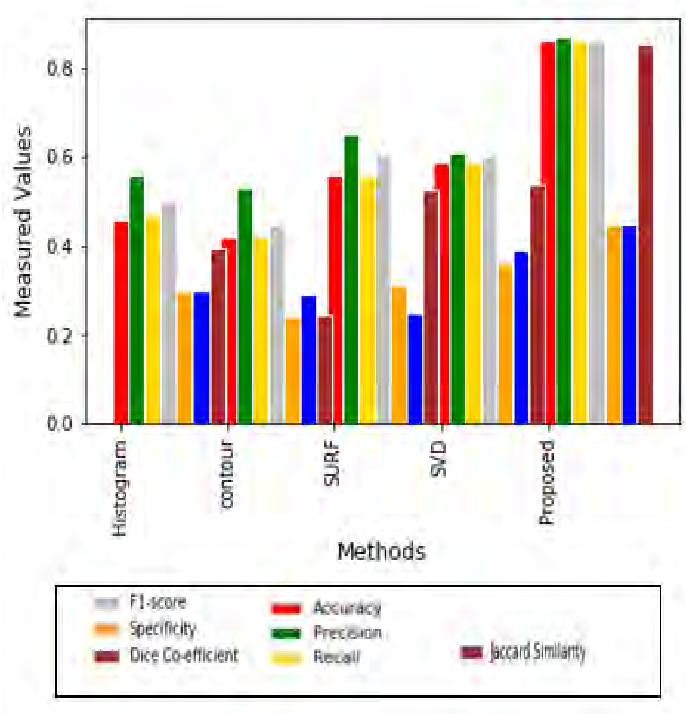


Figure 4.18: Graphical representation of Logo detection for without enhancement.

Table 4.3: Logo extraction with enhancement from document images. Different measuring parameters used for logo extraction.

Logo with Enhancement	Methods	AMBE	Entropy	Michelson contrast	SSIM	PSNR
	Original	0.0	2.3557	1.0	1.0	0
	LPF	6.5397	1.9495	13.941	0.6290	15.376
Image 1	HPF	16.448	1.6693	1.0000	0.2107	2.0304
	GPF	6.5367	1.9514	13.941	0.6290	15.374
	BLF	3.1065	2.8839	1.0	0.9536	19.077
	Original	0.0	2.5487	1.0	1.0	0
	LPF	11.431	2.6607	1.0	0.6263	13.463
Image 2	HPF	13.358	1.6736	1.0	0.1797	3.0717
	GPF	11.424	2.6632	1.0	0.6263	13.461
	BLF	11.052	2.9998	1.0	0.6306	13.528
	Original	0.0	6.70536	1.0	1.0	0
	LPF	34.825	5.0940	6.9375	0.5277	13.553
Image 3	HPF	11.987	1.7308	1.0	0.1834	3.4706
	GF	34.9138	5.10222	6.9375	0.5278	13.543
	BLF	1.4125	6.6202	1.0	0.9410	20.572
	Original	0.0	4.4644	1.0	1.0	0
	LPF	14.2876	3.8093	1.0	0.2015	11.614
Image 4	HPF	13.749	2.3849	1.0	0.0950	2.6578
	GF	14.252	3.8198	1.0	0.2016	11.607
	BLF	2.8882	5.4542	1.0	0.8385	15.910

Logo with Enhancement is given in Figure 4.19 to 4.22.

Figure 4.19: (a) Bilingual document with one Logo. (b) Logo detected.

Figure 4.20: (a). Bilingual document with university prospect logo image. (b). Logo detected.

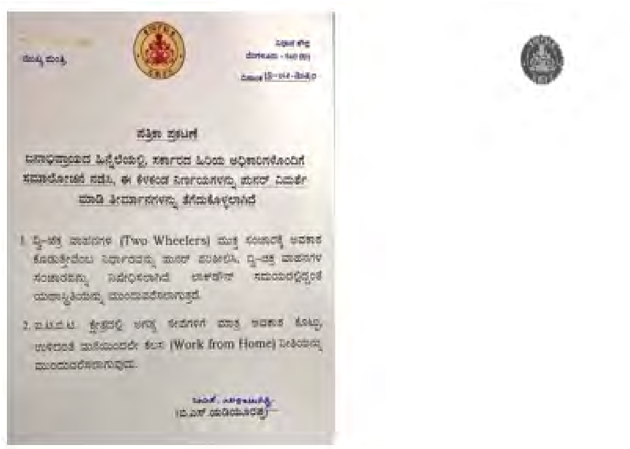


Figure 4.21: (a). Bilingual document with Karnataka Govt. office logo image. (b). Logo detected.

Figure 4.22: (a). Bilingual document with university logo input image. (b). Logo detected.

Figure 4.23: Graphical representation of logo extraction for different enhancement approaches.

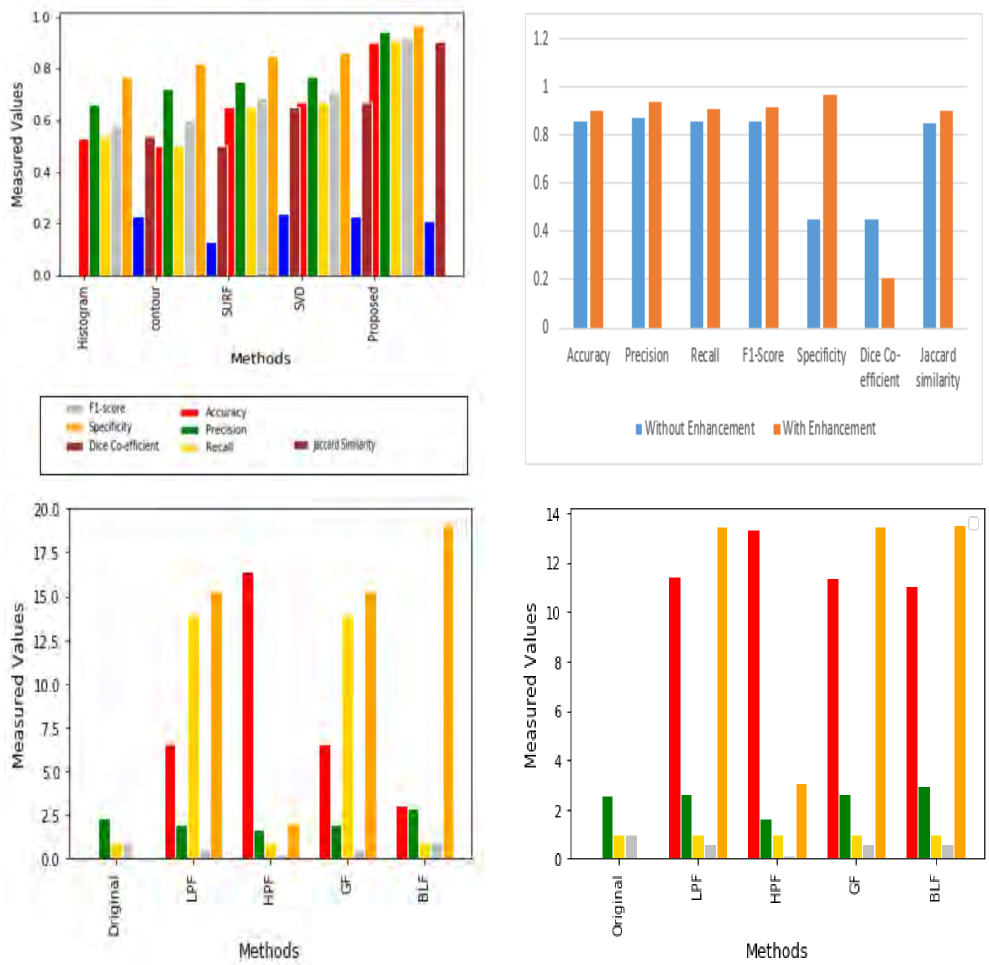


Table 4.4: Different measuring parameters used for logo extraction with enhancement.

Methods	Accuracy	Recall	F1-score	Precision	Specificity	Dice Co-efficient	Jaccard similarity
Histogram	0.53	0.66	0.54	0.58	0.77	0.23	0.537
Contour	0.50	0.72	0.50	0.60	0.82	0.13	0.500
Surf	0.65	0.75	0.65	0.69	0.85	0.24	0.651
SVD	0.67	0.77	0.67	0.71	0.86	0.29	0.670
Proposed	**0.90**	**0.94**	**0.91**	**0.92**	**0.97**	**0.21**	**0.905**

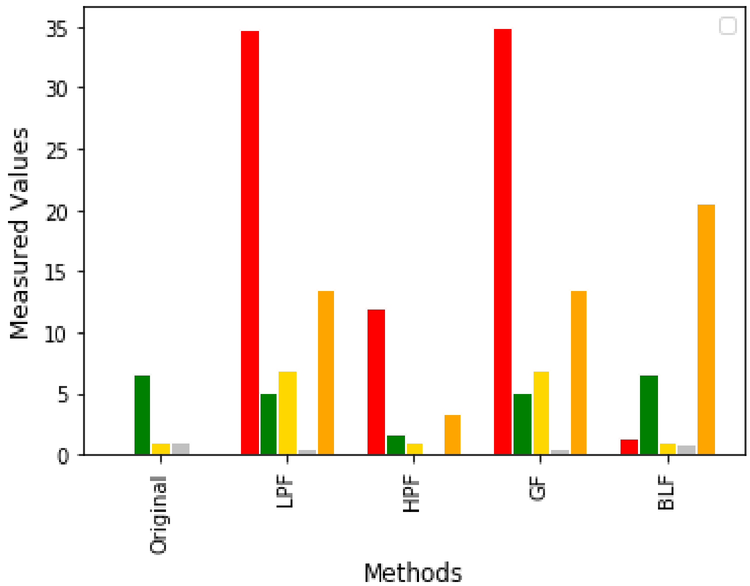


Figure 4.24: Graphical representation for bilingual without enhancement.

Table 4.5: Different measuring parameters used for logo extraction without enhancement and with enhancement

Methods	**Accuracy**	**Recall**	**F1-score**	**Precision**	**Specificity**	**Dice Co-efficient**	**Jaccard similarity**
Without Enhancement	0.86	0.87	0.86	0.86	0.45	0.45	0.851
With Enhancement	0.90	0.94	0.91	0.92	0.97	0.21	0.905

The proposed method gives good result for Accuracy, Precision, Recall, F1-Score, Specificity, Dice Co-efficient, and Jaccard similarity, Further we need to improve the enhancement methods. From the Table 4.5, it is observed that the proposed method gives good results for logo extraction.

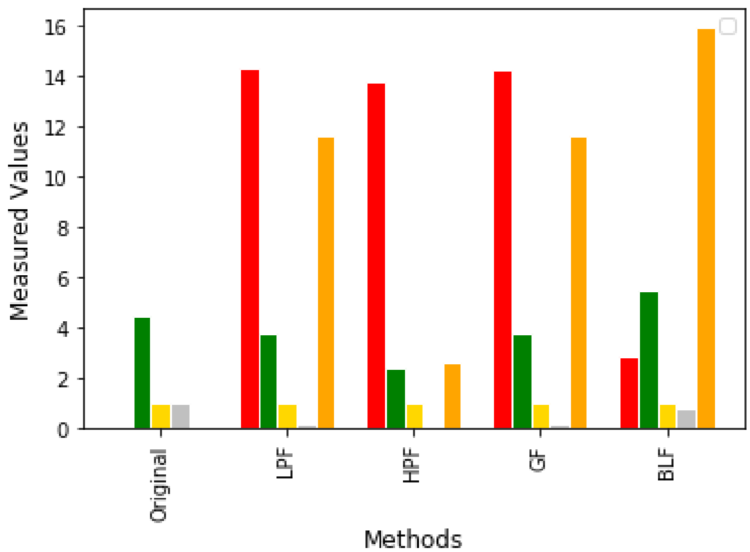


Figure 4.25: Graphical representation for bilingual logo with enhancement and without enhancement document image extraction for different methods.

4.6 Conclusion

In this work, we have presented an approach for signature extraction from a bilingual document images. The proposed approach is based on contour and blob's method. Experimental results show the effectiveness of the proposed method. Experimentation is carried out on real time dataset containing 150 document images. For all the parameters, proposed method gives good results. And also we have presented an efficient algorithm for extraction of logo from bilingual real time document images. The proposed algorithm is tested an two different cases i.e., before enhancement and after enhancement. Logo extraction is done by using masking and median filter techniques. From the experimentation, it is observed for enhancement good result is 90% achived when captured is for without enhancement 86%.

Chapter 5

Segmentation and Skew Estimation at Word Level from Document Images

5.1 Preamble

Section A

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

Some parts of the materials in this chapter have appeared in the following research paper.

1. Ravikumar M, Shivaprasad B.J, Shivakumar G, and Rachana P G. 2019. "Estimation of Skew Angle from Trilingual Handwritten Documents at Word Level: An Approach Based on Region Props". Advances in Intelligent Systems and Computing, 419-426. (Springer Nature Singapore).
2. Shivakumar G, Ravikumar M Sampathkumar S, and Shivaprasad B J. 2022. "Segmentation of Non-Text from Bilingual Real-Time Office Document Images Using U-Net Architecture". The Seybold Report Journal (TSRJ), 17(07), 811–827. (Scopus Indexed).

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.

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 (Sakila et al., 2017). Handwritten documents with multiple orientations require additional pre-processing for segmentation subsequent phases for proper functioning in the handwriting recognition system (Pramanik et al., 2021).

A deep learning approach to detection Distorted scan document angles with different spellings language (Akhter et al., 2020). RLSA algorithm is used Rows and columns of document images (Salagar et al., 2020). 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 (Oliveira et al., 2018). 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 (Ma et al., 2018).

Using FFT (Fast Fourier transform) median filtering, the author proposed skew detection and correction method (Watts et al., 2014). skew detection and correction for Mushaf Al-Quran image pages based totally on Hough rework technique (Bafjaish et al., 2018). 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 (Mandal et al., 2018; Boukharouba et al., 2017).

In handwritten Kannada documents with no constraints, a deskewing algorithm leads to line and word segmentation (Shakunthala et al., 2017). 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) (Khatatneh et al., 2015; Shakunthala et al., 2021). Using a geometrical method, this paper shapes a line from the components separated for various reasons (Soora et al., 2018). A convolutional neural network (CNN) is applied to create the U-Net architecture. An algorithm for segmenting text content lines based on deep learning (Mechi et al., 2019; Saiyed et al., 2021). Using an Adaptive U-Net Architecture for Text Line Segmentation (Gurav et al., 2019).

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 (Deivalakshmi et al., 2013). In this paper, the author uses Symlet wavelet and 2-suggest classification for text segmentation from image documents (Gauttam et al., 2013). 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 (Wang et al., 2021; Ibrahim et al., 2008; Chen et al., 2018; Jobin et al., 2017). Deep Neural Networks (DNN) (Bures et al., 2019). This research utilizes the concept of semantic segmentation with the aid of a multi-scale convolutional neural network community (Dutta et al., 2021). First broadly recognized structure based on Convolutional Neural Networks (CNN) that ultimately cause the wave of research on ANNs this is nevertheless taking area these days (Lombardi et al., 2020). 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 (Minaee et al., 2021).

For better understanding, the remaining document part is organized as follows: proposed methodology is detailed in section 5.2 , followed by result and discussions in section 5.3, finally conclusion is given in section 5.7.

5.2 Proposed Method

In this section, we discuss our approach in detail. Figure 5.1 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 Preprocessing, Data augmentation and Segmentation.

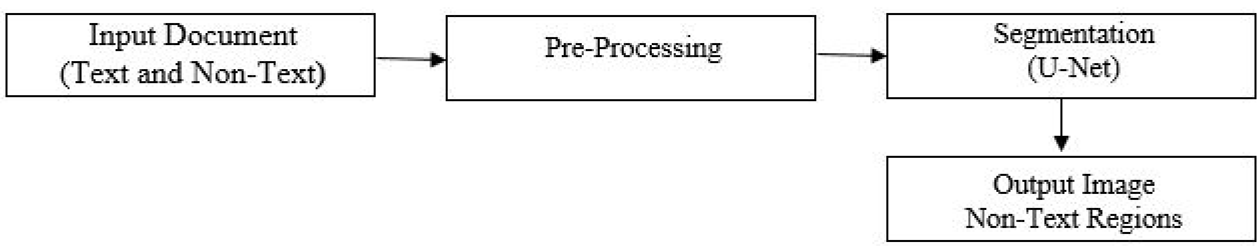


Figure 5.1: Block diagram of the proposed method.

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.

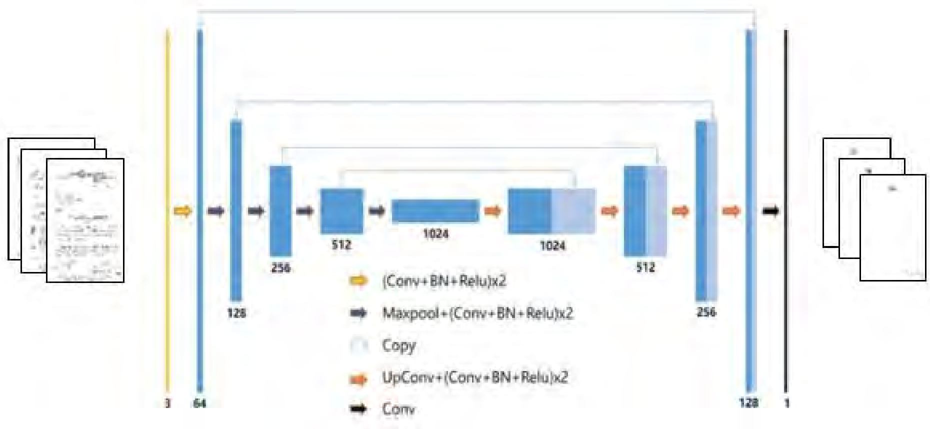


Figure 5.2: U-Net architecture.

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

5.2.1 Preprocessing

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 area 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 encumbrance 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 5.1.

$$\text{Wij} = \frac{(P_{ij}(x,y))^{\gamma}}{\sum_{i=1}^{n} \sum_{j=1}^{m} (P_{ij}(x,y))^{\gamma}} \tag{5.1}$$

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 built as a sum of the transformed W_{ij} weights with degree of membership ψ_{ij}. The enhanced image is given by equation 5.2.

$$\Psi\text{enh(f)} = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} X \Psi_{ij}(f) \tag{5.2}$$

where, ψ_{ij} (f) is image (f) before enhancement. ψ_{ij} (f) is image (f) after enhancement.

5.2.2 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. Figure 5.3 illustrate the principle of the Data augmentation model.

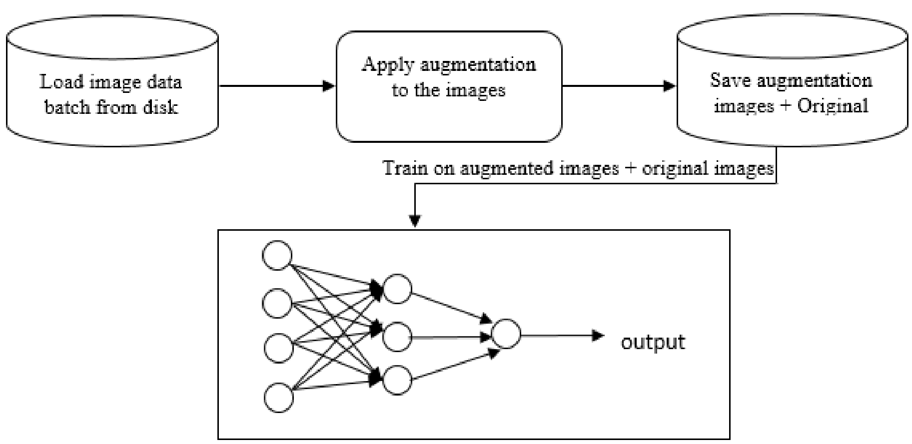


Figure 5.3: Training a deep neural network on both augmented

With a rich and sufficient dataset, the model is better able to estimate results more accurately and efficiently. Figure 5.3 shows the working principle of Data augmentation model.

5.2.3 Segmentation

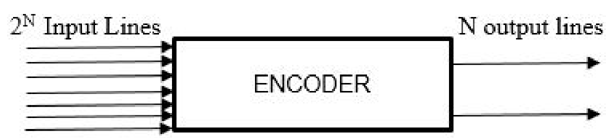
U-Net segmentation uses document images in a CNN (convolutional neural network) structure for picture segmentation. This is correct and green. Data have outperformed an earlier satisfactory approach (a sliding-window convolutional network) for segmenting axonal frameworks in electron microscopic layers. 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. Figure 5.2 shows an architecture with encoders and decoders.

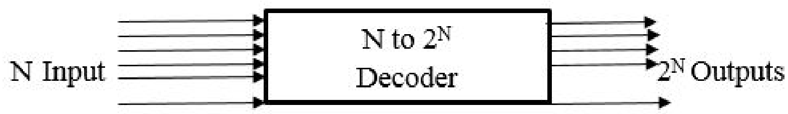
5.2.4 Encoder path:

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 2N input traces. The encoder converts a signal into coded binary output when it receives an input signal.



5.2.5 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.



5.3 Results and Discussion

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. 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 5.4.

(a) (b) (c)

Figure 5.4: shows the resultant output images 1-6, 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.

$$F1Score, F1 = 2 * (P_C * R_C)/(P_C + R_C)(1) \tag{5.3}$$

$$Precision, Pc = (truepositives)/(truepositives + falsepositives) \tag{5.4}$$

$$Sensitivity(Recall), Rc = (truepositives)/(truepositives + falsenegatives) \tag{5.5}$$

$$Specificity = (truenegative)/(truenegative + falsepositives) \tag{5.6}$$

Table 5.1: Comparison of Different quantitative methods.

FPR	**TPR**
0.0	0.2
0.1	0.5
0.2	0.9
0.3	1.0
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

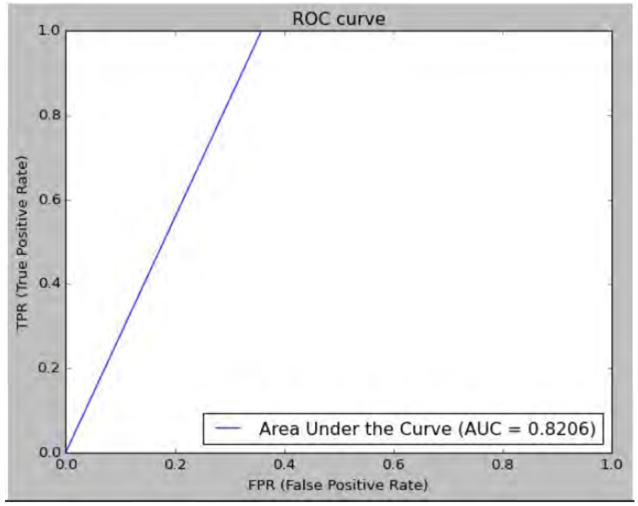


Figure 5.5: ROC diagrams of the proposed U-Net for Real-time dataset.

The values of FPR and TPR for segmentation are given in Table 5.1, and the ROC diagram of the proposed method is plotted in Figure 5.5. Precision and Recall diagram for the proposed method is plotted in Figure 5.6. The accuracy diagram of the proposed method is plotted in Figure 5.7, and the final diagram is plotted in Figure 5.8.

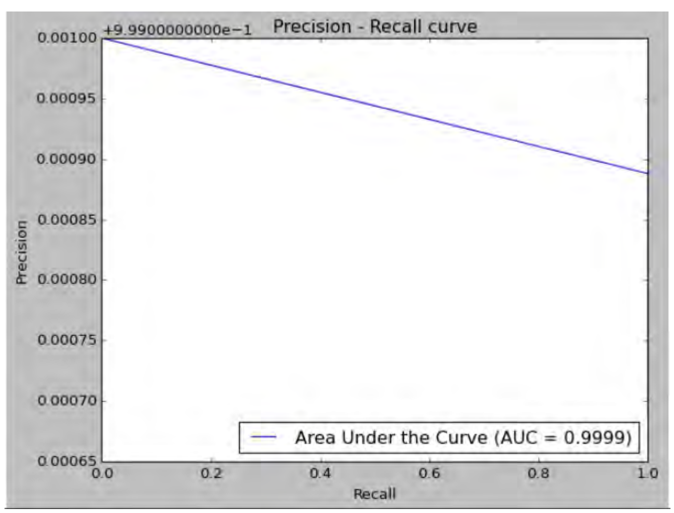


Figure 5.6: Precision and Recall diagram of the present work U-net Segmentation.

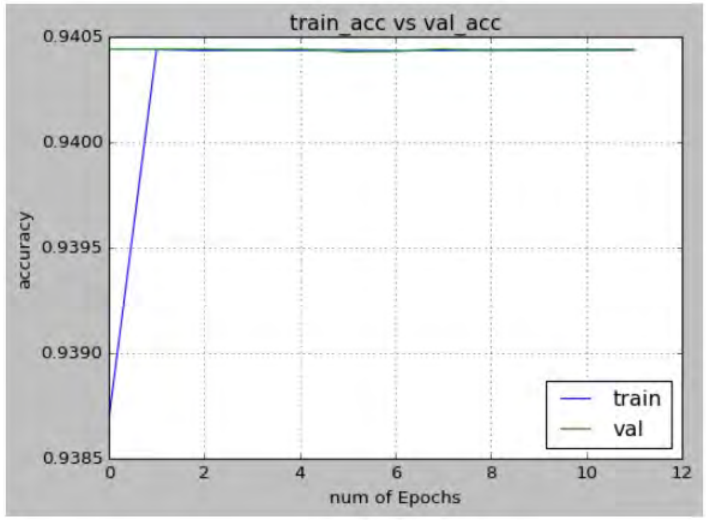


Figure 5.7: Accuracy diagram for the proposed method.

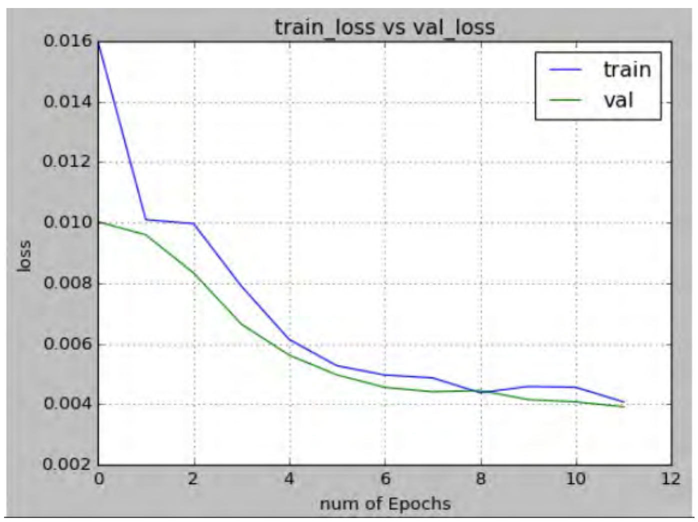


Figure 5.8: Loss diagram for the proposed method.

Different segmentation accuracy methods are compared with proposed method and the values are given in table and graphical representation is shown in Figure 5.9.

Table 5.2: Segmentation Accuracy of different methods.

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

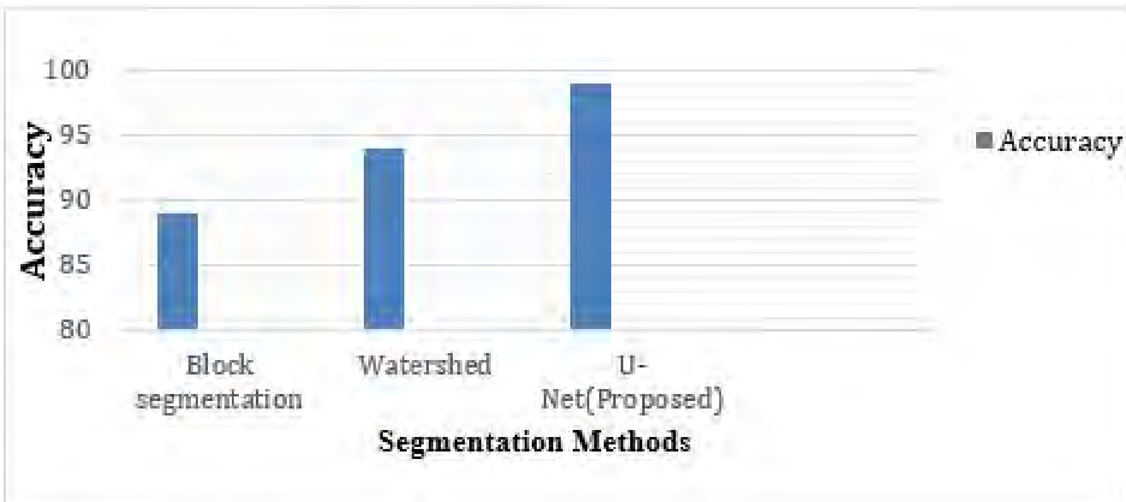


Figure 5.9: Graphical representation of accuracy comparison of different methods.

Finally, different performance measure for segmentation of the proposed method are given in Table 5.3 and its equivalent graphical representation plotted in Figure 5.10

Table 5.3: Segmentation Accuracy of Different measuring parameter methods.

Segmentation Methods	**Accuracy**	**Specificity**	**Sensitivity**	**Precision**	**F1-Score**
Block segmentation	89	59	82	83	82
Watershed	94	60	90	90	90
U-Net(Proposed)	**99**	**64**	**99**	**99**	**99**

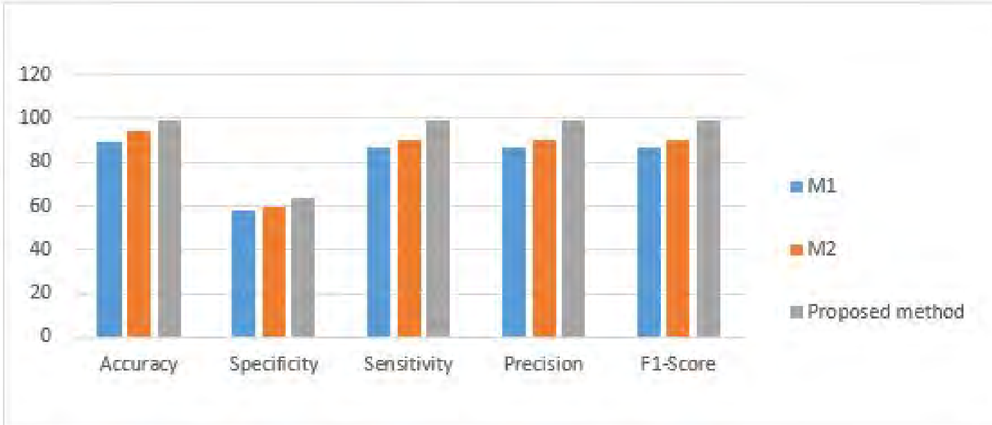


Figure 5.10: Graphical representation of performance comparison for different methods.

5.4 Section B

In document image processing Skew must be estimated and corrected or else analysis of handwritten document becomes difficult and it drastically reduces the recognition rate. Estimation of skew angle can be at block level, line level, and also at word level (Guru et al., 2013). Most of the work has been done on monolingual documents, but very few works have been done on skew estimation and correction for multilingual documents. Reason is difficulty faced while finding the skew angle of multiple words, which are of multiple languages and are at different orientations. Besides its complexities, estimation and correction of multiple skews from multilingual handwritten documents has wide range of applications which include analysis of real-time office documents, analysis of ancient script, and analysis of documents prescribed by the doctors. Remaining part of the paper is divided into three sections. Section 5.5 focuses on proposed methodology in detail, experimentation with results are discussed in Section 5.6, and finally, conclusion is given in Section 5.7.

Next section discusses proposed methodology in detail.

5.5 Proposed Methodology

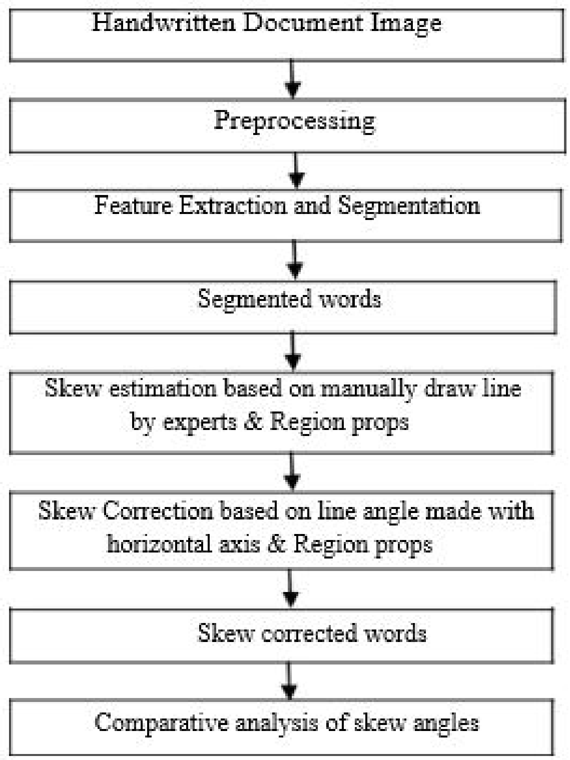


Figure 5.11: Block diagram of the proposed method.

After this, next step is to estimate and correct the skew angle, which can be done using two approaches, and finally, the two approaches are compared for calculating error rates.

In first approach, the skew angle is estimated based on manually drawn lines by human experts, where human experts considered are of different age-group and they are unaware of languages present in the multilingual handwritten documents. They were asked to click manually at two endpoints on the ellipse constructed over word. In order to estimate skewangle, a line is drawn using the clicked points. These clicked points are considered as coordinate values (x1, y1) and (x2, y2) which are stored in the form of an array of two dimensions for further calculations, which are shown in Table 5.4. Then the angle of deviation from the horizontal axis is calculated by the formula which is given below:

Table 5.4: Points on coordinate values of the words.

Point 1		Point 2	
X_1	Y_1	X_2	Y_2
21.50	594.50	269.50	458.50
213.50	1170.50	633.50	1166.50
313.50	450.50	545.50	310.50
349.50	2122.50	637.50	1910.50
605.50	298.50	777.50	166.50
721.50	1850.50	929.50	1714.50
709.50	1138.50	1205.50	1126.50
821.50	154.50	1061.50	−1.50
993.50	1662.50	1453.50	1306.50
1273.50	1134.50	1437.50	1110.50
1517.50	1094.50	2133.50	1090.50
1569.50	186.50	1737.50	390.50
1765.50	470.50	1825.50	574.50
1857.50	630.50	1973.50	810.50
2037.50	858.50	2081.50	926.50

$$\begin{aligned} h &= \sqrt{(x2-x1)^2+(y2-y1)^2} \\ \theta &= \tan^{-1}\left(\frac{y2-y1}{x2-x1}\right) \end{aligned} \tag{5.7}$$

where, h = Hypotenuse. Finally, the estimated skew must be corrected, where skew correction is based on angles of lines drawn manually by experts, which is given in Eq. 5.8.

$$\begin{aligned} x2 &= x1 + h\cos(\theta) \\ y2 &= y1 + h\sin(\theta) \end{aligned} \tag{5.8}$$

In second approach, the skew angle is estimated using the measuring properties of an image region, which is corrected by using geometric transformation function with the help of centroid, major axis, and minor axis. The below formula is used for rotation by an angle θ:

$$x = u\cos(\theta) - v\sin(\theta)$$
$$y = u\sin(\theta) + v\cos(\theta) \tag{5.9}$$

where u and v represent lengths of the major and minor axes. The below formula is also used for rotation by an angle θ in matrix format:

$$R = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{5.10}$$

Finally, the comparative analysis of skew correction among two approaches has been done.

5.6 Experimentation and Results

In order to compare the error rates of two approaches explained above, experimentation is carried out on our own dataset with 300 unconstrained handwritten documents. Once the document is obtained, preprocessing step is applied on the document. The words are labeled in order to avoid the confusion, which also helps in reconstruction after the skew is corrected. We considered five human experts and they are asked to click manually at two end points on the ellipse constructed over word. The below figures shows the different stages of skew estimation and correction.

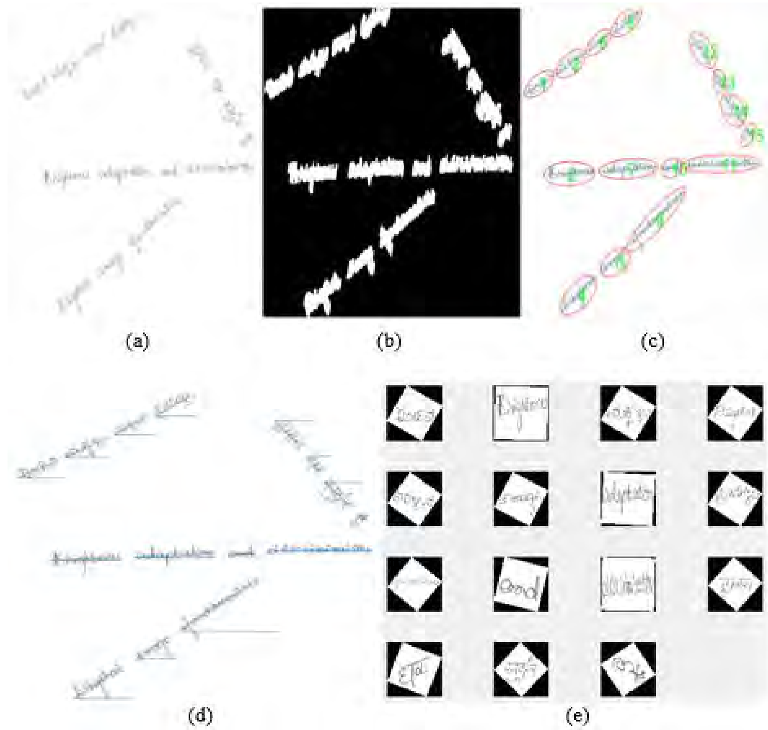


Figure 5.12: (a) Original input image, (b) image after CCA, (c) labeling and line dilation drawn based on expert clicks, and (d) angle found using line drawn by (e) word segmentation and skew correction experts

Figure 5.12(a) shows the original image before any processing, Figure 5.12(b) shows the output of morphological dilation and CCA after noise removal, Figure 5.12(c) is output obtained after human experts clicking two end points and line is drawn on that, it also represents labeling of words. After obtaining a line the angle of deviation from the horizontal is represented in Figure 5.12(d), finally skew corrected words are shown in Figure 5.12(e).

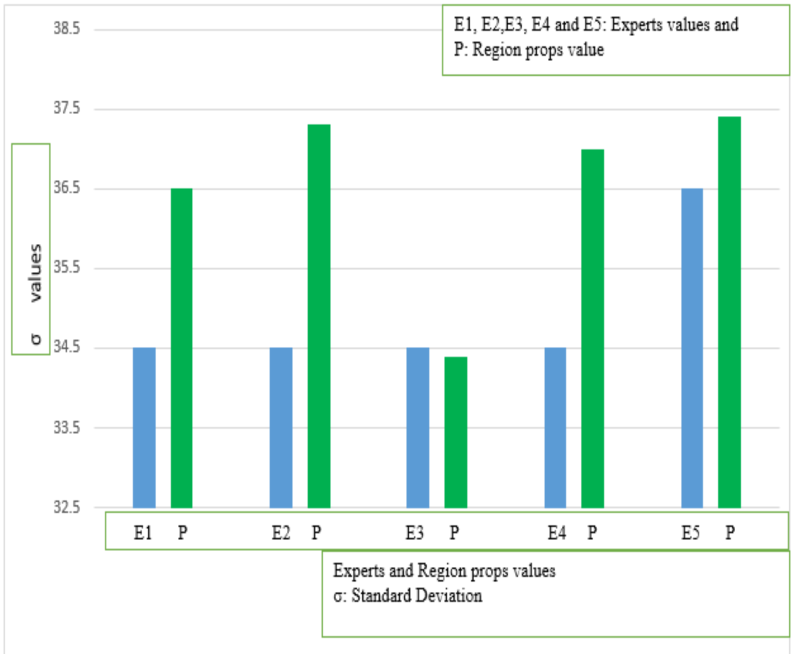


Figure 5.13: Comparison between skew angle of experts and region props.

Figure 5.12(a) shows the original image before any processing, Figure 5.12(b) shows the output of morphological dilation and CCA after noise removal, and Figure 5.12(c) is output obtained after human experts clicking two endpoints and line is drawn on that; it also represents labeling of words. After obtaining a line, the angle of deviation from the horizontal is represented in Figure 5.12(d), and finally skew-corrected words are shown in Figure 5.12(e). Sometimes the error rate of first approach is better compared with the second approach, because experts click the points on character which have low intensity values, and it represents minor axis, but automated system will consider high-intensity values of character which represents major axis.

Figures 5.12(c) and 5.12(d) show the comparative analysis of skew angle among experts and region props, where estimation of skew angle varies and result obtained is higher error rate in human experts when compared with region props. Table 5.5 shows skew angle values of region props and experts.

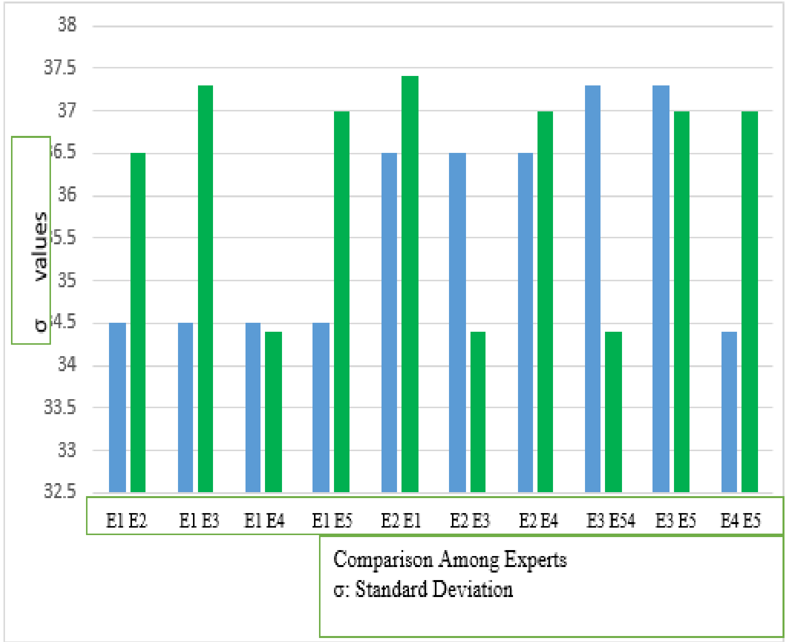


Figure 5.14: Comparison between skew angle among Experts.

Table 5.5: Represents the angle of experts and region props angle.

Expert1 Angle	Expert2 Angle	Expert3 Angle	Expert4 Angle	Expert5 Angle	Region-props Angle
27.29	33.43	28.96	29.62	30.96	29.81
2.14	5.71	5.09	1.09	7.74	1.74
29.53	37.20	33.23	38.15	37.72	31.31
38.36	31.73	35.31	34.77	29.89	36.45
34.24	32.85	33.40	37.69	37.97	35.97
37.40	37.79	34.18	43.60	29.85	27.98
3.751	1.43	2.46	3.25	2.70	3.50
34.18	37.36	30.06	37.72	30.51	33.28
37.24	41.32	37.24	38.77	37.33	39.38
4.085	7.69	11.30	11.00	18.82	13.67
0.73	1.10	1.53	−0.38	−0.75	1.35
−48.90	−50.90	−54.56	−49.39	−48.99	−48.03
−53.13	−65.55	−66.57	−59.30	−60.46	−68.87
−47.07	−51.66	−53.32	−40.60	−53.97	−45.93
−48.36	−38.65	−53.13	−29.74	−57.99	55.97

5.7 Conclusion

In this chapter, U-net architecture to segment non-textual information from a bilingual office document images. Experimentation is carried out on our own dataset and the results shows that performance of the proposed method from the results shows that performance of the proposed method from the results we obtained accuracy of 99%, Specificity-64%, Sensitivity-99%, Precision-99% and F1-Score-99%. And also, an efficient method is proposed for estimation and correction of multiple skews from unconstrained multilingual handwritten documents at the word level, based on the region props. The proposed method is compared with the skew angle, which is obtained from five human experts, and finally, the result is obtained. The results shows that proposed method is efficient than human experts.

Chapter 6

Classification of Text and Non-Text from Bilingual Document Images

6.1 Preamble

An imperative aspect of computer vision is the selection and classification of areas of interest in scanned images of text documents. Many researchers around the world are studying how to convert document images into editable formats. There needs to be a separation of text zones from non-text zones and a correct ordering of them in reading systems. An image can be analyzed to detect/extract/recognize text. For applications including optical character extraction, human-machine input distinction, spam detection, and machine-to-human input differentiation, text recognition and classification in natural images are very significant. Changes in the environment in which images are taken make it difficult for in-text recognition to recognize valuable full text in images. Image text detection identifies locations that contain meaningful whole text in an image. Taking an image in a different area makes it difficult. In analyzing document layouts, it is important to separate text and non-text elements. The complex structure of the document has

Some parts of the materials in this chapter have appeared in the following research paper.

1. Ravikumar M, Shivakumar G, Shivaprasad B J "Classification of Text and Non-Text from Bilingual Real-Time Documents Using Deep Learning Approach," International Journal on Document Analysis and Recognition (IJDAR). (Communicated).

limited the quality of separation results despite several approaches. In order for the printed text to be recognized, it must be separated from non-text areas, such as signatures, handwritten text, logos, and other symbols, in order to be accurate. Most research, however, focuses on converting images of documents into the editable text because of the many ways in which this conversion can be used. Survey of text/non-text separation using various feature classifier combinations [(Ghosh et al., 2018), (Chaithanya et al., 2019),(Tran et al., 2015), (Arvind et al., 2006),(Ghosh et al., 2018), (Puri et al., 2016), (He et al., 2019), (Lee et al., 2018), (Mishra et al., 2018)].

We present an end-to-end deep learning-based framework, called Visual Structure Object Recognition (VSOR), for detecting visual objects in document images, such as tables, figures, and equations. Data-driven and independent of any heuristic rules for detecting visual objects in document images, our framework is based on recent object detection algorithms in computer vision (Chen et al., 2007), (Liu et al., 2021). As our task does not include labeled training data, deep learning-based methods require large amounts of data. VSOR explores transfer learning and domain adaptation in order to solve the scarcity of labeled training data in document images for Visual Structure Object detection/Recognition. The VSOR more accurately localizes all visual objects in document images than state-of-the-art techniques based on numerous public benchmark data sets.

An image description framework based on neural networks. The CNN for image encoding can be replaced with an RNN encoding for the source text, based on the encoder-decoder model used in machine translation (He et al., 2019). Using KCRAlexNet and KCRGoogLeNet (Lee et al., 2018), the structure of this describes the creation of a CNN network, starting from the organization of test data and ending with plotting the test accuracy curve. A visual demonstration of how VSOR technique can successfully locate various visual objects in abstract images within a document. In particular, the contributions of this work are as follows: The aim of this paper is to present an end-to-end trainable deep learning approach based on the concept of object detection algorithms used in recent years in computer vision to locate visual objects (Chen et al., 2007; Puri et al., 2016). To detect visual objects in document images, we refine a model trained by transfer learning. Based on the public standard dataset, our VSOR framework achieves outstanding results.

For better understanding, the remaining document part is organized as follows: Proposed methodology is detailed in section 6.2, followed by result and discussions in section 6.3, finally conclusion is given in section 6.4.

6.2 Proposed Method

In this section, we discuss our approach in detail. Figure 6.1 shows a flowchart of the proposed method. The proposed method mainly consists of three different stages they are pre-processing, Data augmentation, classification 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, tables, equations, graph & 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 Classification.

In the subsequent sections, we discuss the Preprocessing, Region proposal network, Data augmentation and Classification.

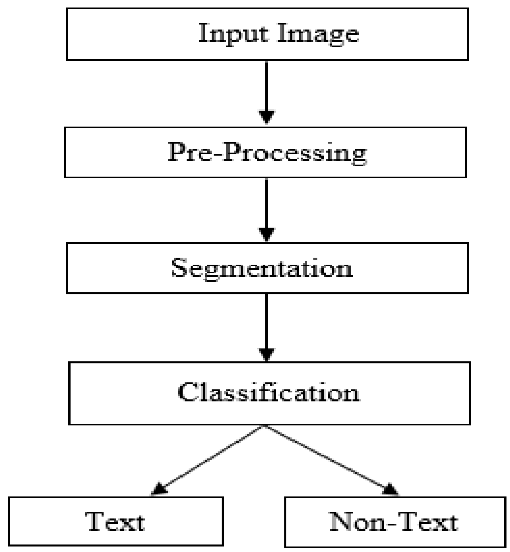


Figure 6.1: The Architecture of the Proposed Method.

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 and Components-Based Region method is used.

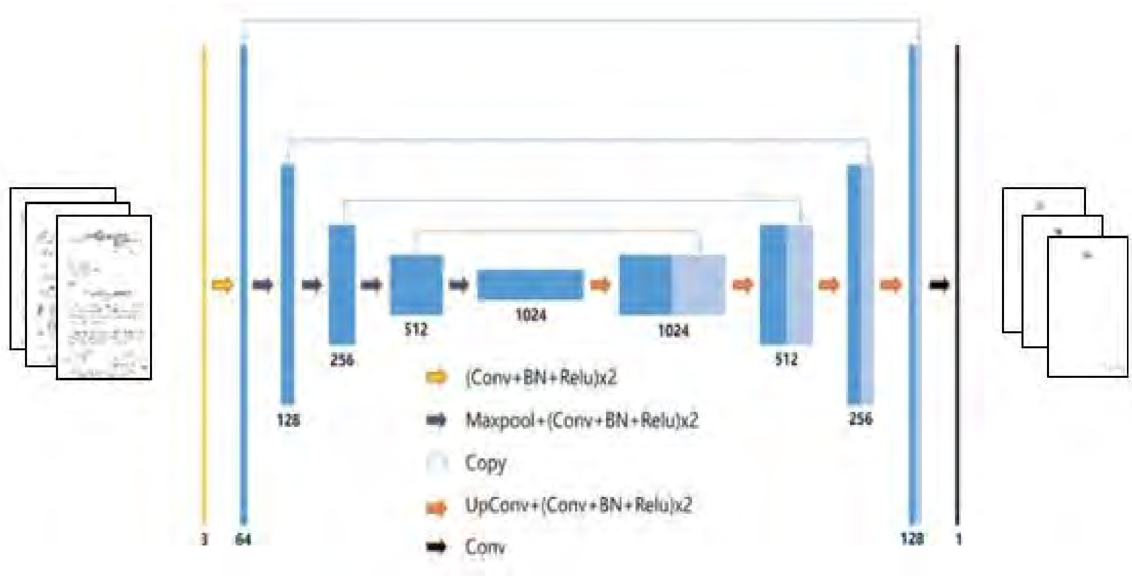


Figure 6.2: U-Net architecture.

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

6.2.1 Data acquisition

A dataset of 10000 real-time document images was collected from publicly different places or regions dataset. All these images have 2480X3508 to 1242X1754 pixels of size and converted gray-scale for processing further, Figure 6.2 shows the text and non-text images.

In order to conduct experiments, we collected and annotated a dataset for document page object detection. Throughout the dataset, we have selected images from each of the 2500 classes of English and Kannada documents. Each document page object is manually annotated; there are 10000 objects in total. As far as page objects are concerned, documents are indeed unbalanced. The networks are trained with 80% of the data, and they are tested with 20%.

6.2.2 Preprocessing

After data axquisition images are fed into preprocessing, to improve the quality of an image, because the intensity value of real-time images varies on the imaging camera and scanner used. Hence, intensity is normalized to reduce the bias in image. Normalized inner and inter class combinations for enhanced using Spatial domain methods, Frequency domain methods (DFT) and Fuzzy approach, better enhancement methods gives good qualitative results.

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.

6.2.3 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.

6.2.4 Classification

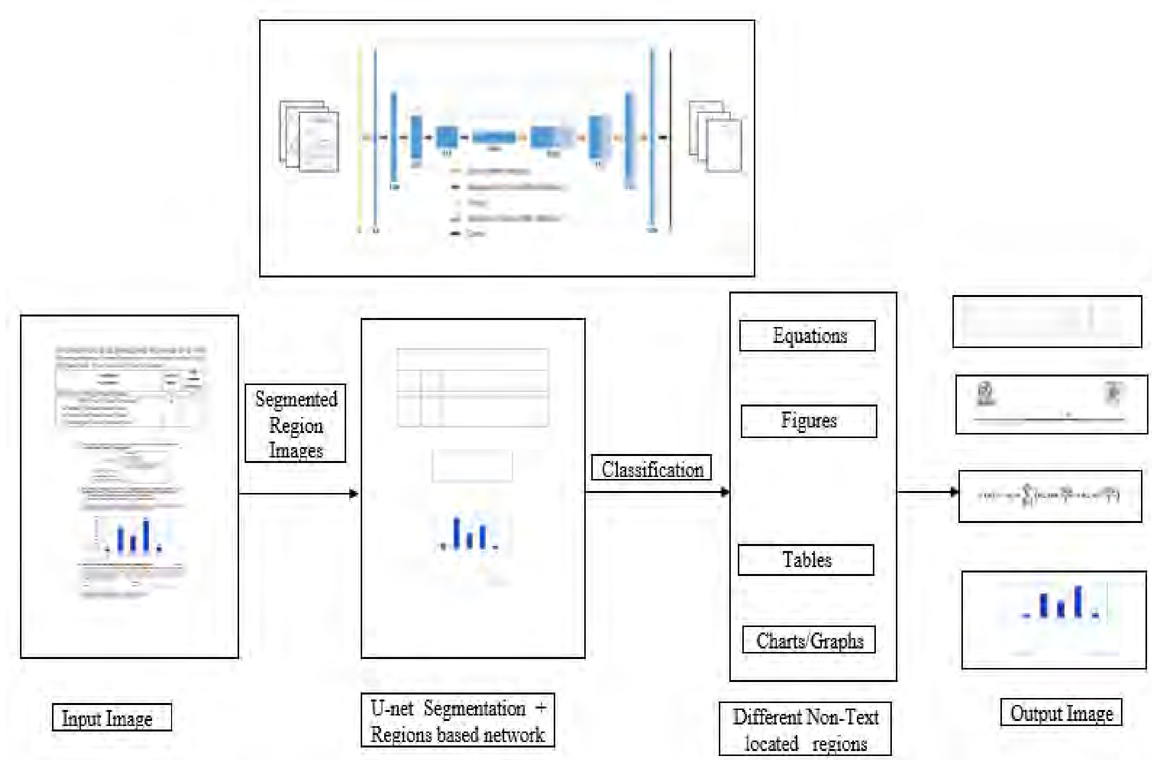


Figure 6.3: Frame work for the classification of methods.

In this section, we discuss our approach in detail. Figure 6.3 shows a frame work of the classification method. Our proposed approach is compared with state-of-the-art deep learning models using VGG-16 pre-trained on ImageNet and BERT base pre-trained on document images. The BERT base was fine-tuned with class labels after tokens were extracted from document images. The VGG-16 model was directly used for classifying document images based on pre-trained features. There is no further fine-tuning to this process. Our model was compared to U-net segmentation, which simplifies BERT models with knowledge distillation and is designed explicitly for document-level classification.

As we embed the U-net into the component-based region, the original bounding box refinement still remains. The remaining will be used to segment the next document.

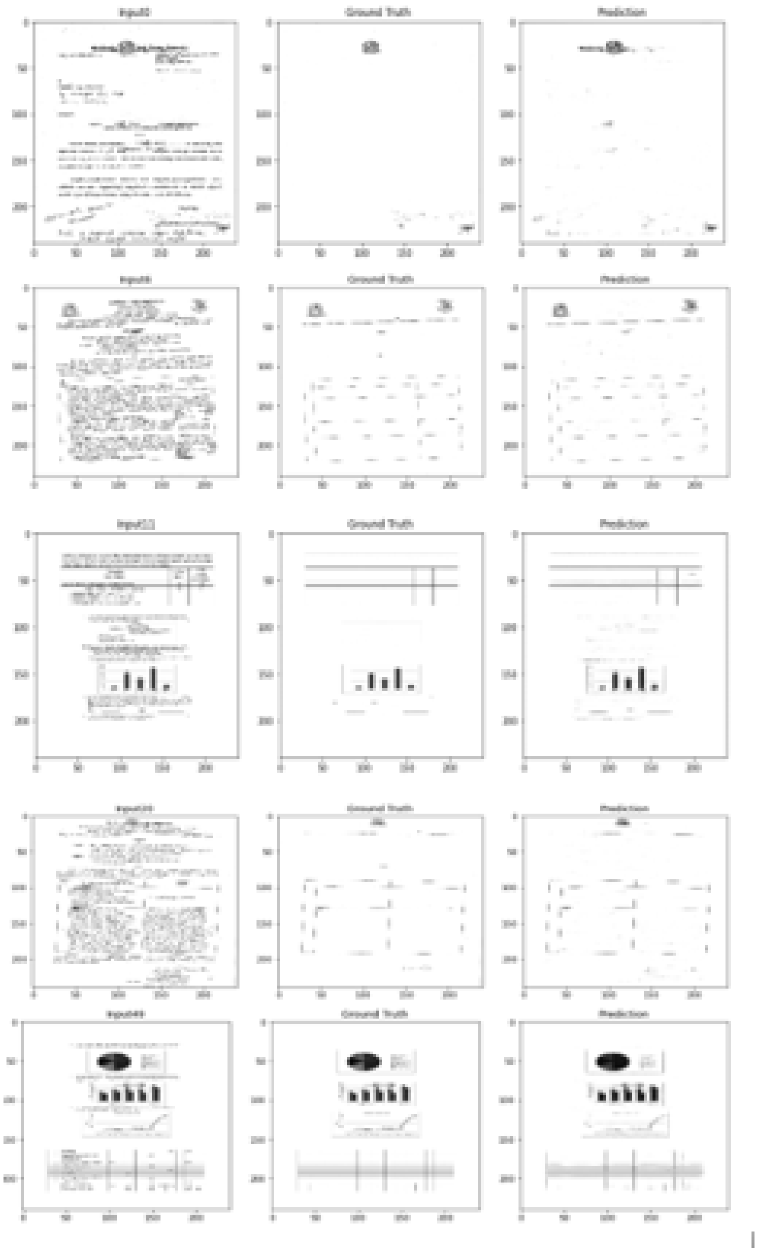
This method applies the component-based region as a generative method to detect and localize an image. Our results for document image detection and localization are significantly improved when the bounding box of the image is added to the U-net structure, as demonstrated in printed real-time document image experiments. Additionally, we can use our U-net neural network to detect equations, graphs, tables, charts, and segment images with minimal adjustment.

Combining the two techniques does improve the sample AUC by for Eff-GNN + Word2Vec and by 86.0 % for U-net segmentation and Region Component based on the Tobacco-3482 dataset. Using Word2Vec text features alone does not show much improvement over U-net segmentation and Region Component based classification on in real-time document image datasets. There may be enough information in printed document images to classify both textual and image content. In addition, the BERT and VGG-16 models achieve similar classification results.

In Table 6.2, the classification Accuracy of 10000 real-time naturally captured camera and scanned document data sets are compared. As compared to models in BERT families (93.95 %) and VGG-16 (93.16 %), Our proposed model achieves 99.0 % under real-time data sets for Office Documents, Advertisement boards, Inauguration boards, Direction boards, and Answer scripts. Moreover, we tested whether could combine text and image embedding in the U-net and component region-based approach.

6.3 Results and Discussion

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. 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 6.4.



(a) (b) (c)

Figure 6.4: shows the resultant output images 1-5, 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.

$$F1Score, F1 = 2 * (P_C * R_C)/(P_C + R_C)(1) \tag{6.1}$$

$$Precision, Pc = (truepositives)/(truepositives + falsepositives) \tag{6.2}$$

$$Sensitivity(Recall), Rc = (truepositives)/(truepositives + falsenegatives) \tag{6.3}$$

$$Specificity = (truenegative)/(truenegative + falsepositives) \tag{6.4}$$

Table 6.1: Comparison of Different quantitative methods.

FPR	**TPR**
0.0	0.2
0.1	0.5
0.2	0.9
0.3	1.0
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

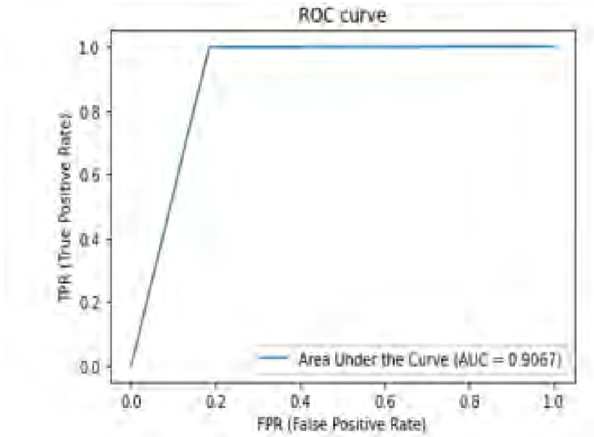


Figure 6.5: ROC diagrams of the proposed U-Net for Real-time dataset.

The values of FPR and TPR for segmentation are given in Table 6.1, and the ROC diagram of the proposed method is plotted in Figure 6.5 Precision and Recall diagram for the proposed method is plotted in Figure 6.6. The accuracy diagram of the proposed method is plotted in Figure 6.7, and the final diagram is plotted in Figure 6.8.

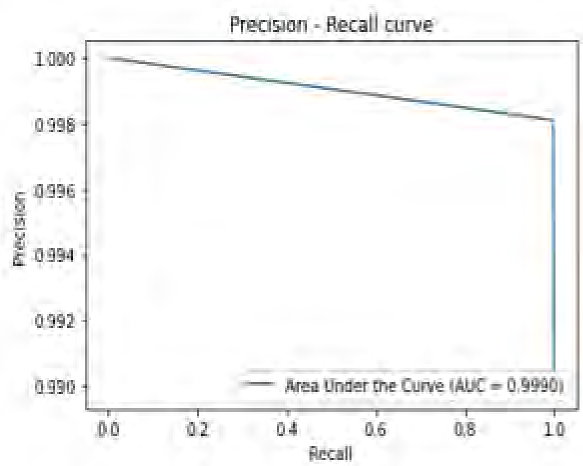


Figure 6.6: Precision and Recall diagram of the present work U-net Segmentation.

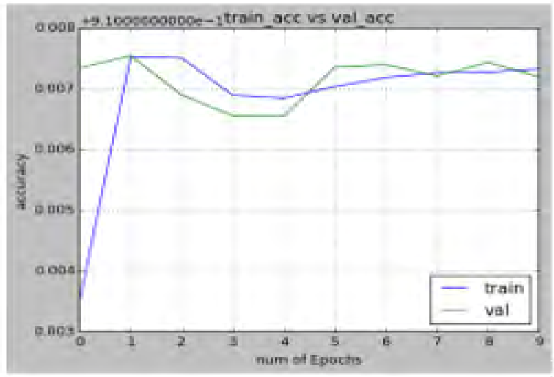


Figure 6.7: Accuracy diagram for the proposed method.

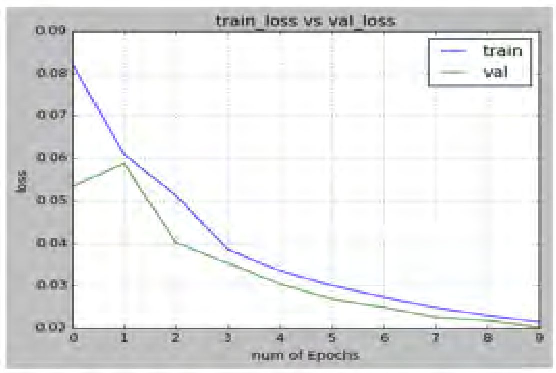


Figure 6.8: Loss diagram for the proposed method.

Different classification accuracy methods are compared with proposed method and the values are given in table and graphical representation is shown in Figure 6.9.

Table 6.2: Classification Accuracy of different methods.

Classification Methods	**Accuracy**
VGG-16	93.16
DocBert	90.18
BERT	93.95
Eff-GNN + Word2Vec	86.0
Our (Proposed)	**99.62**

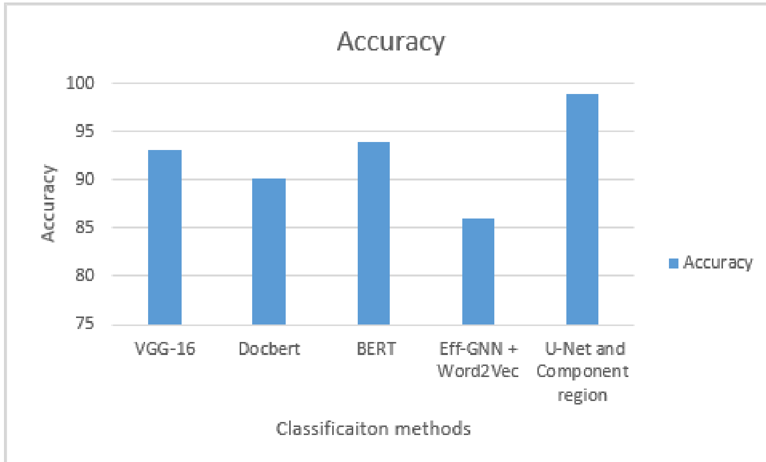


Figure 6.9: Graphical representation of accuracy comparison of different methods.

Finally, different performance measure for classification of the proposed method are given in Table 6.3 and its equivalent graphical representation plotted in Figure 6.10.

Table 6.3: Classification Accuracy of Different measuring parameter methods.

Classification Methods	**Acc-uracy**	**Speci-ficity**	**Sensi-tivity**	**Prec-ision**	**F1-Score**
VGG-16	93	80	89	83	86
Docbert	90	82	88	89	88
BERT	93	86	87	91	92
Eff-GNN+Word2Vec	86	68	78	75	79
U-Net (Proposed)	**99**	**81**	**99**	**99**	**99**

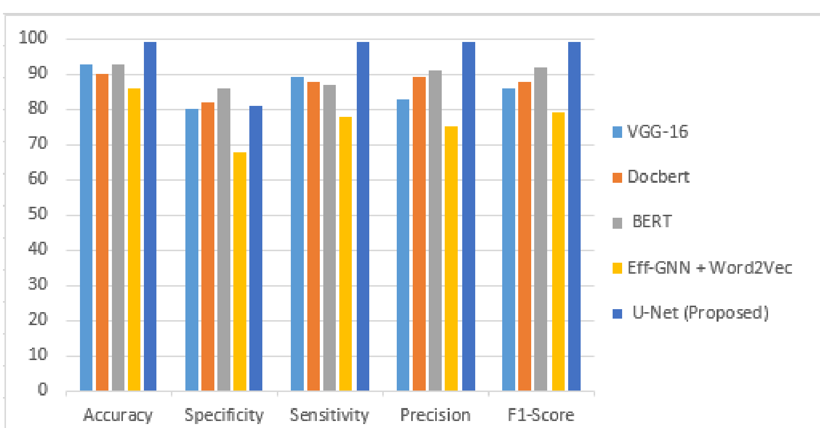


Figure 6.10: Graphical representation of performance comparison for different methods.

6.4 Conclusion

As deep learning has shown good performance in detecting natural scene objects, the paper explores how to employ this technique for detecting document page objects (like tables, formulae, equations, and figures). A deep neural network model is modified and improved based on the differences between document images and natural images. By combining U-Net Segmentation and component-based region proposal, we propose a method for generating moderate candidate regions for objects with multi-scale problems. Additionally, this paper examines the significant factors affecting the performance of deep neural networks for detecting page objects, including network structure and training strategies. As a result of experimenting on our own dataset, the performance of the proposed method was demonstrated as 99% accuracy, 81% specificity, 99% sensitivity, 99% precision, and 99% F1-Score obtained from the results.

Chapter 7

Epilogue

7.1 Preamble

The separation of text and non-text printed/handwritten document images poses many challenges in the field of document image analysis, especially when there is bilingual text, blurry or distorted images, or shadows on the documents. It is inevitable that distortions will appear in printed/handwritten documents, but separating text from non-text poses a challenge. For this reason, an efficient approach to extracting text and non-text from printed, handwritten, or scanned text needs to be developed. Consequently, deciphering bilingual document images into text and non-text can be divided into two broad categories: statistical strategies (local strategies) and texture-based strategies (global strategies). Lines, words, and characters are used for local segmentation in the LWC method. Logos, tables, equations, signatures, lines, words, and characters must first be segmented before components are available. The models we provide can also be used to separate text from non-text information in document images. A novel method was also devised to identify images in bilingual printed and handwritten documents and to distinguish them from the text.

In this chapter, all the model proposals are summarized. Further, the major contributions of the research work are listed. Subsequently, the scope for further research in this direction is also highlighted.

7.2 Summary

Chapter 1, gives introduction of Text and Non-Text printed/handwritten bilingual document image enhancement, segmentation and Classification. The state of art existed methods and brief survey methods are presented.

In Chapter 2, We just made an attempt to brief our effort towards creation of the datasets. The first dataset consists of printed/handwritten bilingual document images with multiple skews and logos, pictures, tables, equations, numbers, seal impressions.

In chapter 3, We have proposed an efficient approach for enhancement of real time document images. The proposed approach Fuzzy Logic(FL) approach perform better than the existing methods. To enhance the real time document images, the different spatial domain enhancement methods are used like Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Gamma Correction(GC), Linear Transformation(LT), Log Transform (LogT) and Contrast Stretching(CS). Frequency domain methods focus on the image orthogonal transform instead of the image itself. Low Pass Filter(LPF), High Pass Filter (HPF), Gaussian Filter (GF), Non Local Means (NLM), Constrained Least Squares (CLS), Pseudo Inverse Filter(PIF) are used.

In chapter 4, We have presented an approach for signature extraction from a bilingual document images. The proposed approach is based on contour and blob's method. The proposed algorithm is tested a two different cases i.e., before enhancement and after enhancement. Logo extraction is done by using masking and median filter techniques. To measure the performance different performance metrics are used like, Accuracy, Precision, Recall and F1-score. The proposed method is compared with existing methods; our method performs well.

In chapter 5, We proposed a hybrid U-SegNet model which integrates both U-net and SegNet architectures. The performance is evaluated using metrics like accuracy, precision, recall and F1-score and comparison analysis is also conducted with other segmentation methods such as watershed method, Fuzzy C- means and U-net method.

In chapter 6, We have proposed U-net and component-based region network is a different method for analyzing features of document images, such as regions, bounding boxes, convex hulls, filters, and enhancements, when compared with existing methods. The performance metrics used to measure performance are Accuracy, Precision, Recall, and F1-Score. The proposed method is compared with existing methods. Our method performs well.

In brief, following are the major contributions of the research work presented in this thesis work.

7.3 Contributions

A successful attempt towards creation of considerably large datasets viz., A dataset consisting of 10000 bilingual document images of five different document each with 2000 images.

- To proposed an efficient classifier for classification of text and non-text information.
- To develop an effective segmentation algorithm for extracting the text.
- To design an algorithm for estimation and correction of skew angle.

7.4 Scope for Future Work

Text and non-text separation is an important processing step in any document analysis system. It then divides off-line printed/handwritten document images into several types based on the nature of the problems each finds, in an attempt to provide understanding of the various techniques presented in the literature.

To the best of our knowledge, no work has been documented on extracting text and non-text extraction from printed/handwritten bilingual document images with several challenges. As a result, in our research, we propose developing a novel contour approach and applying bounding boxes to various sections of document images. The developed model can be employed with or without a priori knowledge of document images for both handwritten and printed documents, and it can extract various text and non-text document images if they exist. The designed model is intended to recognize and extract document images if they exist, and it can be used with or without prior knowledge of text and non-text separation of document images for both handwritten and printed documents. The proposed models efficacy will be tested on a large number of document images with various parameters.

Author's Publications

➢ Journals

1. Shivakumar G, Ravikumar M Sampathkumar S, and Shivaprasad B J. (2022). "Segmentation of Non-Text from Bilingual Real-Time Office Document Images Using U-Net Architecture", The Seybold Report Journal (TSRJ), 17(07), 811–827. (Scopus Indexed).

2. Ravikumar M, Shivakumar G and Shivaprasad B J. (2022). "Enhancement of Real Time Document Images Using Fuzzy Logic and Machine Learning Approach", Journal of Jilin University (Engineering and Technology Edition Vol:41 Issue:09:2022. (Scopus Indexed).

3. Ravikumar M, Shivakumar G and Shivaprasad B J "Classification of Text and Non-Text from Bilingual Real-Time Documents Using Deep Learning Approach", International Journal on Document Analysis and Recognition. (Communicated).

➢ Conferences

1. Ravikumar M., Shivaprasad B.J, Shivakumar G, and Rachana P G. 2019. "Estimation of Skew Angle from Trilingual Handwritten Documents at Word Level: An Approach Based on Region Props", Advances in Intelligent Systems and Computing, 419-426. (Springer).

2. Ravikumar M., and Shivakumar G 2020. "A Survey on Text Detection from Document Images", In International Conference on Intelligent Computing and Smart Communication. Springer, Singapore, pp. 961-972. (springer).

3. Shivakumar G., Ravikumar M., Shivaprasad B. J., and Guru D. S., 2022. "Signature Extraction from Bilingual Document Images Using Blobs Method", In Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough. Springer, Cham, pp. 283-294. (springer).

4. Shivakumar G., Ravikumar M., Shivaprasad B J and Guru D. S., 2022. "Extraction of Logo from Real Time Document Images Using Masking and Median Filter Approaches", IEEE INCET, Technically Co-Sponsored by IEEE Bangalore Section and IEEE USA, pp. 01-07. (IEEE).

Bibliography

Abdulmunim, M. E., and Abass, H. K. (2017). Logo matching in Arabic documents using region based features and SURF descriptor. In Annual Conference on New Trends in Information and Communications Technology Applications (NTICT), IEEE, pages 75-79.

Agrawal, P., and Varma, R. (2012). Text Extraction from Images. Text Extraction from Images, IJCSET, Vol. 2, Issue 4, pages 1083-1087.

Ahmed, S., Malik, M. I., Liwicki, M., and Dengel, A. (2012). Signature segmentation from document images. In 2012 International Conference on Frontiers in Handwriting Recognition, pages 425-429.

Akhter, S. S. M. N., and Rege, P. P. (2020). Improving skew detection and correction in different document images using a deep learning approach. 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pages 01–06.

Alaei, A., and Delalandre, M. (2014). A Complete Logo Detection/Recognition System for Document Images. 11th IAPR International Workshop on Document Analysis Systems, pages 01-05.

Alaei, A., Delalandre, M., and Girard, N. (2013). Logo Detection Using Painting Based Representation and Probability Features. 12th International Conference on Document Analysis and Recognition, pages 1235-1239.

Ali Bagheri, M., and Gao, Q. (2012). Logo recognition based on a novel pairwise classification approach. In The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), IEEE, pages 316-321.

Al-Khatatneh, A., Pitchay, S. A., and Al-qudah, M. (2015). A review of skew detection techniques for document, 17th UKSim-AMSS International Conference on Modelling and Simulation (UKSim), IEEE, pages 316-321.

Ambica Rani., and Harinderpal Singh Er. (2015). A Review On Various Techniques for Skew Detection and Correctin In Handwritten Text Documents. IJREAT

International Journal of Research in Engineering and Advanced Technology, Vol. 3, Issue 3, pages 58-67.

Arief, R., Mutiara, A. B., and Kusuma, T. M. (2018). Automated extraction of large scale scanned document images using Google vision OCR in apache Hadoop environment. International Journal of Advanced Computer Science and Applications, Vol. 9, No. 11, pages 112-116.

Arvind, K. R., Pati, P. B., and Ramakrishnan, A. G. (2006). Automatic text block separation in document images. In Fourth International Conference on Intelligent Sensing and Information Processing, IEEE, pages 53-58.

Babu, D. R., Kumat, P. M., and Dhannawat, M. D. (2006). Skew angle estimation and correction of handwritten, textual and large areas of non- textual document images: A novel approach. International conference on image processing, computer vision and pattern recognition, pages 01-07.

Bafjaish, S. S., Azmi, M. S., Al-Mhiqani, M. N., Radzid, A. R., and Mahdin, H. B. (2018). Skew detection and correction of mushaf al-quran script using hough transform. International Journal of Advanced Computer Science and Applications, Vol. 9, No. 8, pages 402-409.

Balamurugan, E., Sengottuvelam, P., and Sangeetha, K. (2013). Performance Analysis on Point Operations Based Image Enhancement for Document Images. International journal of engineering research and technology, Vol. 2, Issue 11, pages 955-961.

Bandyopadhyay, S. K., Bhattacharyya, D., and Das, P. (2007). Handwritten Signature Extraction from Watermarked Images using Genetic Crossover. International Conference on Multimedia and Ubiquitous Engineering, pages 987-991.

Banerjee, P., and Chaudhuri, B. B. (2012). A system for handwritten and machine-printed text separation in Bangla document images. In 2012 International Conference on Frontiers in Handwriting Recognition, IEEE, pages 758-762.

Banka, R., and Nourbakhsh, F. (2004). Extraction of Signature and Handwritten Regions from Official Binary Document Images. 17th International Conference on Pattern Recognition, pages 01-04.

Bannigidad, P., and Gudada, C. (2016). Restoration of Degraded Historical Kannada Handwritten Document Images Using Image Enhancement Techniques. International Conference on Soft Computing and Pattern Recognition, pages 498-508.

Bavdekar, S. B. (2015). Using tables and graphs for reporting data. The Journal of the Association of Physicians of India, Vol. 63, Issue 10, pages 59-63.

Baviskar, D., Ahirrao, S., Potdar, V., and Kotecha, K. (2021). Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions. IEEE Access, pages 01-46.

Bedil, S. S., and Khandelwal, R. (2013). Various Image Enhancement Techniques - A Critical Review. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, pages 1605-1609.

Bezmaternykh, P. V., and Nikolaev, D. P. (2020). A document skew detection method using fast Hough transform. International Conference on Machine Vision, pages 01-06.

Bezmaternykh, P. V., Ilin, D., and Nikolaev, D. P. (2019). U-net-bin: hacking the document image binarization contest, Computer Optics, Vol. 43, Issue 5, pages 825–832.

Bhavani, A., and Kumar, B. S. (2021). A Review of State Art of Text Classification Algorithms. In 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, pages 1484-1490.

Bhosle, V. V., and Pawar, V. P. (2017). Automatic Logo Extraction and Detection for Document Verification using SIFT and SURF, International Journal of Engineering Research and Technology (IJERT), Vol. 6, pages 555-560.

Bhowmik, S., Sarkar, R., Nasipuri, M., and Doermann, D. (2018). Text and non-text separation in offline document images: a survey, International Journal on Document Analysis and Recognition (IJDAR), Vol. 21, Issue 1, pages 01-20.

Biswas, S., Kim, T. H., and Bhattacharyya, D. (2010). Features extraction and verification of signature image using clustering technique. International Journal of Smart Home, Vol. 4, Issue 3, pages 43-55.

Blessie, E. C., and Deepa, A. (2019). Classification of Text Documents Using Adaptive Robust Classifier. International Journal of Recent Technology and Engineering, Vol. 7, Issue 6, pages 1482-1489.

Boukharouba, A. (2017). A new algorithm for skew correction and baseline detection based on the randomized Hough Transform. Journal of King Saud University – Computer and Information Sciences, Vol. 29, Issue 1, pages 01–10.

Brindha, B., and Anusuya, P. K. (2015). Image Enhancement Techniques: A Review, International Journal of Research in Engineering and Technology, Vol. 4, Issue 5, pages 455-459.

Brodi, D., and Milivojevi, Z. N. (2012). Estimation of the Handwritten Text Skew Based on Binary Moments. Radio engineering, Vol. 21, pages 162-169.

Buades, A., Coll, B., and Morel, J. M. (2005). A non-local algorithm for image denoising, In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE, Vol. 2, pages 60-65.

Bures, L., Gruber, I., Neduchal, P., Hlavac, M., and Hruz, M. (2019). Semantic text segmentation from synthetic images of full-text documents, pages 1380– 1405.

Byun, H. R., Roh, M. C., Kim, K. C., Choi, Y. W., and Lee, S. W. (2002, August). Scene text extraction in complex images. In International Workshop on Document Analysis Systems, Springer, Berlin, Heidelberg, pages 329-340.

Chaithanya, C. P., Manohar, N., and Issac, A. B. (2019). Automatic Text Detection and Classification in Natural Images, International Journal of Recent Technology and Engineering, Vol. 7, Issue-5S3, pages 176-180.

Chen, J., and Lopresti, D. P. (2011). Table Detection in Noisy Off-line Handwritten Documents. International Conference on Document Analysis and Recognition, pages 399-403.

Chen, J., Leung, M. K., and Gao, Y. (2003). Noisy logo recognition using line segment Hausdorff distance. Pattern Recognit, Vol. 36, pages 943-955.

Chen, J., Shao H., and Hu, C. (2017) Image segmentation based on mathematical morphological operator, pages 23–41.

Chen, N., and Blostein, D. (2007). A survey of document image classification: problem statement, classifier architecture and performance evaluation. International Journal of Document Analysis and Recognition (IJDAR), Vol. 10, Issue 1, pages 01-16.

Chethana, H. T., and Mamatha, H. R. (2016). Comparative study of text line segmentation on handwritten kannada documents, International Journal of Computer Science and Information Technologies, Vol. 7, Issue 1, pages 26–33.

Chidiac, N. M., Damien, P., and Yaacoub, C. (2016). A robust algorithm for text extraction from images. In 39th International Conference on Telecommunications and Signal Processing (TSP), IEEE, pages 493-497.

Cho, H., Wang, J., and Lee, S., (2012). Text image deblurring using text-specific properties. In European Conference on Computer Vision, Springer, Berlin, Heidelberg, pages 524-537.

Chowdhury, S. P., Dhar, S., Das, A. K., Chanda, B., and McMenemy, K. (2009). Robust Extraction of Text from Camera Images, ICDAR, IEEE, pages 1280-1284.

Cortes, C., Fisher, K., Pregibon, D., and Rogers, A. (2000). Hancock: A Language for Extracting Signatures from Data Streams, Proceedings of the sixth ACM International conference on Knowledge discovery and data mining, pages 09-17.

Cüceloğlu, İ., and Oğul, H. (2014). Detecting handwritten signatures in scanned documents. In Proceedings of the 19th computer vision winter workshop, pages 89-94.

Deivalakshmi, S., Palanisamy, P., and Vishwanathan, G. (2013). A novel method for text and non-text segmentation in document images. International Conference on Communication and Signal Processing, IEEE, pages 255-259.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In IEEE conference on computer vision and pattern recognition, pages 248-255.

Dhandra, B. V., Soma, S., Rashmi, T., and Gururaj, M. (2010). Classification of Document Image Components. International Journal of Engineering Research and Technology, Vol. 2, Issue 10, pages 1429-1439.

Diligenti, M., Frasconi, P., and Gori, M. (2003). Hidden tree Markov models for document image classification. IEEE Transactions on pattern analysis and machine intelligence, Vol. 25, Issue 4, pages 519-523.

Dixit, U. D., and Shirdhonkar, M. S. (2015). Automatic logo extraction from document images, International Journal on Cybernetics and Informatics (IJCI), Vol. 4, Issue 2, 0. pages 221-226.

Dixit, U. D., and Shirdhonkar, M. S. (2016). Automatic logo detection and extraction using singular value decomposition. In International Conference on Communication and Signal Processing (ICCSP), IEEE, pages 787-790.

Dixit, U. D., and Shirdhonkar, M. S. (2017). Signature based Document Image Retrieval Using Multi-level DWT Features. International Journal of Image, Graphics and Signal Processing, Vol. 9, Issue 8, 42, pages 42-49.

Dutta, A., Garai, A., Biswas, S., and Das, A. K. (2021). Segmentation of text lines using multi-scale cnn from warped printed and handwritten document images. Int. J. Document Anal.Recognit. 24, pages 299–313.

Elmannai, W., Elleithy, K. M., and Pande, V. (2012). Efficient and Robust Optical Character Recognition Algorithm for Signature Recognition. 25th International Conference on Computer Applications in Industry and Engineering, pages 01-06.

Elrajubi, O. M., and El-Feghi, I. S. (2015). Angle features extraction of handwritten signatures. In International Conference on Computer Vision and Image Analysis Applications, IEEE, pages 01-04.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, Vol. 88, Issue 2, pages 303-338.

Fang, J., Gao, L., Bai, K., Qiu, R., Tao, X., and Tang, Z. (2011). A table detection method for multipage pdf documents via visual seperators and tabular structures. In 2011 International Conference on Document Analysis and Recognition, IEEE, pages 779-783.

Farahmand, A., Sarrafzadeh, H., and Shanbehzadeh, J. (2013). Document image noises and removal methods. Lecture Notes in Engineering and Computer Science, pages 01-05.

Farooq, F., Sridharan, K., and Govindaraju, V. (2006). Identifying handwritten text in mixed documents. In Proc. International Conference on Pattern Recognition, pages 01-04.

Fehérvári, I., and Appalaraju, S. (2019). Scalable logo recognition using proxies. In Winter Conference on Applications of Computer Vision, IEEE, pages 715-725.

Firdausy, K., Sutikno, T., and Prasetyo, E. (2007). Image enhancement using contrast stretching on RGB and IHS digital image. Indonesian Journal of Electrical Engineering, Vol. 5, Issue 1, pages 45-50.

Fletcher, L. A., and Kasturi, R. (1988). A robust algorithm for text string separation from mixed text/graphics images. IEEE transactions on pattern analysis and machine intelligence, Vol. 10, Issue 6, pages 910-918.

Fourure, D., Emonet, R., Fromont, E., Muselet, D., Tremeau, A., and Wolf, C. (2017). Residual conv-deconv grid network for semantic segmentation. ArXiv abs/1707.07958, pages 01–12.

Ganchimeg, G. (2015). History document image background noise and removal methods, International Journal of Knowledge Content Development and Technology, Vol. 5, Issue 2, pages 11-24.

Gatos, B., Danatsas, D., Pratikakis, I., and Perantonis, S. J. (2005, August). Automatic table detection in document images. In International Conference on Pattern Recognition and Image Analysis, Springer, Berlin, Heidelberg, pages 609-618.

Gautam, A. (2013). Segmentation of text from image document. International journal of computer science and information technologies, Vol. 4, Issue 3, pages 538-540.

Gautam, C. M., Sharma, S., and Verma, J. S. (2012). A GUI for Automatic Extraction of Signature from Image Document. International Journal of Computer Applications, 54, pages 13-19.

Ghosh, R., and Mandal, G. (2012). Skew Detection and Correction of Online Bangla Handwritten Word. International Journal of Computer Science Issues, Vol. 9, Issue 4, No 2, pages 202-205.

Ghosh, R., Mandal, G. (2012). Skew detection and correction of online Bangla handwritten word. IJCSI Int. J. Comput. Sci. Vol. 9, Issues 4, 202.

Ghosh, S., Hassan, S. K., Khan, A. H., Manna, A., Bhowmik, S., and Sarkar, R. (2022). Application of texture-based features for text non-text classification in printed document images with novel feature selection algorithm. Soft Computing, Vol. 26, Issue 2, pages 891-909.

Ghosh, S., Lahiri, D., Bhowmik, S., Kavallieratou, E., and Sarkar, R. (2018). Text/non-text separation from handwritten document images using LBP based features: An empirical study, Journal of Imaging, Vol. 4, Issue 4, 57, pages 01-15.

Gilani, A., Qasim, S. R., Malik, I., and Shafait, F. (2017). Table detection using deep learning. In 2017 14th IAPR international conference on document analysis and recognition (ICDAR), Vol. 1, IEEE, pages 771-776.

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440-1448.

Glagolevs, J., and Freivalds, K. (2017). Logo detection in images using HOG and SIFT. In 2017 5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering, pages 01-05.

Gupta, A., Tiwari, D., Khurana, T., and Das, S. (2019). Table detection and metadata extraction in document images. In Smart Innovations in Communication and Computational Sciences, pages 361-372.

Gurav, A. A., and Nene, M. J. (2019). Word segmentation in document images using deep convolutional encoder decoder network. IEEE 5th International Conference for Convergence in Technology (I2CT), pages 01–06.

Guru, D. S., Ravikumar, M., and Manjunath, S. (2013). Multiple Skew Estimation in Multilingual Handwritten Documents. International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, pages 65-69.

Guru, D. S., Suhil, M., Ravikumar, M., and Manjunath, S. (2015). Small eigenvalue based skew estimation of handwritten devanagari words, in: MIKE, pages 216-225.

Hafemann, L. G., Sabourin, R., and Oliveira, L. (2017). Offline handwritten signature verification — Literature review. 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pages 01-08.

Han, D. (2013). Comparison of commonly used image interpolation methods, In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering, ICCSEE, Vol. 10, pages 1556-1559.

Hao, L., Gao, L., Yi, X., and Tang, Z. (2016). A table detection method for pdf documents based on convolutional neural networks. In 12th IAPR Workshop on Document Analysis Systems (DAS), IEEE, pages 287-292.

Harraj, A. E., and Raissouni, N. (2015). OCR Accuracy Improvement on Document Images Through a Novel Pre-Processing Approach. Signal and Image Processing: An International Journal. 6., pages 01-15.

Hasan, Y. M., and Karam, L. J. (2000). Morphological text extraction from images, IEEE Transactions on Image Processing, Vol. 9, No. 11, pages 1978-1983.

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961-2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778.

He, S., and Lu, Y. (2019). A Modularized Architecture of Multi-Branch Convolutional Neural Network for Image Captioning. Electronics, Vol. 8, Issue 12, 1417, pages 01-15.

Hu, J., Kashi, R., and Wilfong, G. (1999). Document image layout comparison and classification. In Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR, IEEE, pages 285-288.

Huang, K., Chen, Z., Yu, M., Yan, X., and Yin, A. (2019). An efficient document skew detection method using probability model and q test, Electronics, Vol. 9, Issue 1, 55, pages 01–17.

Huang, Y., Wu, R., Sun, Y., Wang, W., and Ding, X. (2015). Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy. IEEE Transactions on Intelligent Transportation Systems, Vol. 16, Issue 4, pages 1951-1960.

Huang, Z., and Leng, J. (2014). Text extraction in natural scenes using region-based method. Journal of Digital Information Management, Vol. 12, pages 246-254.

Ibrahim, Z., Isa, D., and Rajkumar, R. (2008). Text and non-text segmentation and classification from document images. In International Conference on Computer Science and Software Engineering, Vol. 1, IEEE, pages 973-976

Ikonomakis, M., Kotsiantis, S., and Tampakas, V. (2005). Text classification using machine learning techniques. WSEAS transactions on computers, Vol. 4, Issue 8, pages 966-974.

Iwasokun, G. B., and Oluwole, C. A. (2014). Image Enhancement Methods: A Review, British Journal of Mathematics and Computer Science, Vol 4, Issue 16, pages 2251-2277.

Jadhav, T. (2019). Handwritten signature verification using local binary pattern features and KNN. Int. Res. J. Eng. Technol, Vol. 6, Issue 4, pages 579-586.

Jadhav, T. (2019). Handwritten Signature Verification using Local Binary Pattern Features and KNN, International Research Journal of Engineering and Technology (IRJET), Vol. 6 Issue 4, pages 579-586.

Jain, S. A., Rani N. S., and Chandan, N. (2018). Image Enhancement of Complex Document Images Using Histogram of Gradient Features. International Journal of Engineering and Technology, 7 (4.36), pages 780-783.

Javed, M., Nagabhushan, P., and Chaudhuri, B. B. (2014). Direct Processing of Run Length Compressed Document Image for Segmentation and Characterization of a Specified Block. arXiv preprint arXiv:1402.1971., International Journal of Computer Applications, Vol. 83, No. 15, pages 01-08.

Jo, J., Koo, H. I., Soh, J. W., and Cho, N. I. (2020). Handwritten text segmentation via end-to-end learning of convolutional neural networks, Multimedia Tools and Applications, pages 01–14.

Jobin, K.V., and Jawahar, C. V. (2017). Document image segmentation using deep features, in: NCVPRIPG, pages 01–11.

Joshi, S., and Kumar, S. (2018). Image contrast enhancement using fuzzy logic, Computer Vision and Pattern Recognition, pages 01-04.

Julca-Aguilar, F. D., Maia, A. L., and Hirata, N. S. (2017). Text/non-text classification of connected components in document images, In 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE, pages 450-455.

Kamboj, and Poonam. (2014). Image Enhancement with Different Techniques and Aspects, International Journal of Computer Science and Information Technologies, Vol. 5, Issue 3, pages 4301-4303.

Karanje, U. B., and Dagade, R. (2014). Survey on text detection, segmentation and recognition from a natural scene images, International Journal of Computer Applications, Vol. 108, Issue 13, pages 39-43.

Kasar, T., Barlas, P., Adam, S., Chatelain, C., and Paquet, T. (2013). Learning to detect tables in scanned document images using line information. In 2013 12th

International Conference on Document Analysis and Recognition, IEEE, pages 1185-1189.

Kasiviswanathan, H., Ball, G, R., and Srihari, S, N. (2010). Top down analysis of line structure in handwritten documents. 20th International conference on pattern recognition.

Kaur, A. (2018). A review on image enhancement with deep learning approach, ACCENTS Transactions on Image Processing and Computer Vision, Vol. 4, Issue 11, pages 16-20.

Kaur, R., and Taqdir. (2016). Image Enhancement Techniques-A Review, International Research Journal of Engineering and Technology (IRJET), Vol. 3, Issue 3, pages 1308-1315.

Kaur, S., Garg, P., and Sharma, S. (2017). Image Enhancement Techniques Based On Histogram Equalization, International Journal OF Engineering Sciences and Management Research, pages 23-29.

Kavallieratou, E., Fakotakis, N., Kokkinakis, G. (2002). Skew angle estimation for printed and handwritten documents using the Wigner-Ville distribution. ELSEVIER Image Vis. Comput. 20, pages 813–824.

Kavasidis, I., Palazzo, S., Spampinato, C., Pino, C., Giordano, D., Giuffrida, D., and Messina, P. (2018). A saliency-based convolutional neural network for table and chart detection in digitized documents. arXiv preprint arXiv:1804.06236, pages 01-13.

Khalil, R., and Al-Jumaily, A. (2020). Hybridization of Local Search Optimization and Support Vector Machine Algorithms for Classification Problems Enhancement, ICICCT 2019, LAIS 9, pages 209–217.

Konya, I. V., Eickeler S. S., and Seibert C. (2010). Fast seamless skew and orientation detection in document images, 20th International Conference on Pattern Recognition, pages 1924–1928.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., and Brown, D. (2019). Text classification algorithms: A survey. Information, Vol. 10, Issue 4, 150, pages 01-68.

Kuhnke, K., Simoncini, L., and Kovacs-V, Z. M. (1995). A system for machine-written and hand-written character distinction. In Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 2, IEEE, pages 811-814.

Kumar, A. (2014). An Efficient Approach for Text Extraction in Images and Video Frames Using Gabor Filter. International Journal of Computer and Electrical Engineering, Vol. 6, No. 4, pages 02-07.

Kumar, A., Rastogi, P., and Srivastava, P. (2015). Design and FPGA Implementation of DWT Image Text Extraction Technique. Procedia Computer Science, 57, pages 1015-1025.

Kumar, S. S., Rajendran, P., Prabaharan, P., and Soman, K. P. (2016). Text/image region separation for document layout detection of old document images using non-linear diffusion and level set. Procedia Computer Science, 93, pages 469-477.

Le, V. P., Nayef, N., Visani, M., and Ogier, J. M. (2016). Time-efficient Logo Spotting using Text/Non-text Separation as Preprocessing and Approximate Nearest Neighbor Search. In Semaine du Document Numérique et de la Recherched'Information SDNRI 2016 (CORIA-CIFED), pages 365-380.

Le, V. P., Nayef, N., Visani, M., Ogier, J. M., and De Tran, C. (2015). Text and non-text segmentation based on connected component features. 13th International Conference on Document Analysis and Recognition, IEEE, pages 1096-1100.

Lee, S. G., Sung, Y., Kim, Y. G., and Cha, E. Y. (2018). Variations of AlexNet and GoogLeNet to improve Korean character recognition performance. Journal of Information Processing Systems, Vol. 14, Issue 1, pages 205-217.

Li, X., Yin, F., and Liu, C. (2018). Page object detection from pdf document images by deep structured prediction and supervised clustering. 24th International Conference on Pattern Recognition (ICPR), pages 3627-3632.

Li, Z., and Luo, J. (2011). Resolution Enhancement from Document Images for Text Extraction, In 2011 Fifth FTRA International Conference on Multimedia and Ubiquitous Engineering, IEEE, pages 251-256.

Li, Z., Schulte-Austum, M., and Neschen, M., (2010). Fast logo detection and recognition in document images. In 20th International Conference on Pattern Recognition, IEEE, pages 2716-2719.

Lin, H., Yang, F., and Yen, S. (2001). Off-Line Verification for Chinese Signatures. Int. J. Comput. Process. Orient. Lang, 14, pages 17-28.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision. Springer, Cham, pages 740-755.

Lin, Y., Song, Y., Li, Y., Wang, F., and He, K. (2017). Multilingual corpus construction based on printed and handwritten character separation. Multimedia Tools and Applications, Vol.76, Issue 3, pages 4123-4139.

Liu, L., Wang, Z., Qiu, T., Chen, Q., Lu, Y., and Suen, C. Y. (2021). Document image classification: Progress over two decades. Neurocomputing, 453, pages 223-240.

Liu, Z., and Smith, R. (2013). A simple equation region detector for printed document images in tesseract. 12th International Conference on Document Analysis and Recognition, IEEE, pages 245-249.

Lombardi, F., and Marinai, S. (2020). Deep learning for historical document analysis and recognition—a survey, Journal of Imaging, 6, pages 01–30.

Lu, D., Huang, X., and Sui, L. (2018). Binarization of degraded document images based on contrast enhancement, International Journal on Document Analysis and Recognition (IJDAR), Vol. 21, Issue 1, pages 123-135.

Lyu, P., Yao, C., Wu, W., Yan, S., and Bai, X. (2018). Multi-oriented scene text detection via corner localization and region segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7553-7563.

Ma, K., Shu, Z., Bai, X., Wang, J., and Samaras, D. (2018). Document image unwarping via a stacked u-net', IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4700–4709.

Madasu, V. K., and Lovell, B. C. (2008). An automatic off-line signature verification and forgery detection syste, Pattern Recognition Technologies and Applications: Recent Advances. Information Science Reference, pages 63-89.

Magudeeswaran, V., and Ravichandran, C. G. (2013). Fuzzy logic-based histogram equalization for image contrast enhancement, Mathematical problems in engineering, pages 01-10.

Maini, R., and Aggarwal, H. (2010). A Comprehensive Review of Image Enhancement Techniques, Computer Vision and Pattern Recognition, Journal of Computing, Vol. 2, Issue 3, ISSN 2151-9617, pages 08-13.

Malhotra, A., and Kumar, E. D. (2015). Image documentation for the Enhancement of text and noisy images, IJARCCE, Vol. 4, Issue 9, pages 62-65.

Mamatha, B. S., and Chaithra, B. P. (2014). Extraction of text from images, IJECS, Vol. 3, Issue 8, pages 7583-7587.

Mandal, G. (2018). An unprecedented approach of skew detection and correction for online bengali handwritten words, International Journal of Advanced Research in Computer Engineering and Technology, Vol. 7, Issue 2, pages 155-159.

Mandal, R., Roy P. P., and Pal, U. (2012). Signature segmentation from machine printed documents using contextual information, International Journal of Pattern Recognition and Artificial Intelligence, 26(07), 1253003, pages 01-23.

Mandal, R., Roy, P. P., and Pal, U. (2011). Signature segmentation from machine printed documents using conditional random field, In International Conference on Document Analysis and Recognition, IEEE, pages1170-1174.

Mandal, R., Roy, P. P., Pal, U., and Blumenstein, M. (2013). Signature segmentation and recognition from scanned documents, In 13th International Conference on Intelligent Systems Design and Applications, IEEE, pages 80-85.

Mandivarapu, J.K., Bunch, E., You, Q., and Fung, G.M. (2021). Efficient Document Image Classification Using Region-Based Graph Neural Network. ArXiv, abs/2106.13802. pages 01-08.

Manjunath Aradhya, V. N., Basavaraju, H. T., and Guru, D. S. (2021). Decade research on text detection in images/videos: a review. Evolutionary Intelligence, Vol. 14, Issue 2, pages 405-431.

Mazzei, A., Kaplan, F., and Dillenbourg, P. (2010). Extraction and classification of handwritten annotations. In 1st International Workshop on Paper Computing, a Ubicomp (Paper Comp), pages 01-04.

Mechi, O., Mehri, M., Ingold, R., and Amara, N. E. B. (2019) Text line segmentation in historical document images using an adaptive U-Net architecture, In 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, pages 369-374.

Melo, V. K. S. L., and Dantas, B. B. L. (2018). A fully convolutional network for signature segmentation from document images. In 16th International Conference on Frontiers in Handwriting Recognition, pages 540-545.

Minaee, S., and Wang, Y. (2017). Text extraction from texture images using masked signal decomposition, In IEEE Global Conference on Signal and Information Processing (GlobalSIP), pages 1210-1214.

Minaee, S., Boykov, Y., Porikli, F. M., Plaza, A. J., Kehtarnavaz N., and Terzopoulos, D. (2021). Image segmentation using deep learning: A survey, IEEE transactions on pattern analysis and machine intelligence, pages 01–23.

Mishra, D., Malathi, D., and Senthilkumar, K. (2018). Digit Recognition Using Deep Learning, International Journal of Pure and Applied Mathematics, Vol. 118, No. 22, pages 295-302.

Moghaddam, R. F. and Cheriet, M. (2009). Low quality document image modeling and enhancement, International Journal on Document Analysis and Recognition (IJDAR), Vol. 11, Issue 4, pages 183-201.

Moll, M. A., and Baird, H. S. (2008). Segmentation-based retrieval of document images from diverse collections. In Document Recognition and Retrieval XV (Vol. 6815, p. 68150L). International Society for Optics and Photonics, pages 01-09.

Mondal, R., Bhowmik, S., and Sarkar, R. (2020). tsegGAN: a generative adversarial network for segmenting touching nontext components from text ones in handwriting, IEEE Transactions on Instrumentation and Measurement, 70, pages 01-10.

Murty, M. R., Murthy, J. V. R., and PVGD, P. R. (2011). Text document classification based on a least square support vector machines with singular value decomposition. Int. J. Comput. Appl. (IJCA), pages 21-26.

Mustafa, W. A., Yazid, H., and Jaafar, M. (2018). An Improved Sauvola Approach on Document Images Binarization, Journal of Telecommunication, pages 43-50.

Naqvi, S. A., Rehan, W., and Zafar, F. M. (2011). Efficient Logo Extraction Method Using Mountain Function in Gridding Technique. Sindh University Research Journal-SURJ (Science Series), 43(1 (a)), pages 01-05.

Narasimha Reddy, S., Deshpandande, P. S. (2017). A novel local skew correction and segmentation approach for printed multilingual Indian documents, Alexandria Engineering Journal, 2017. Pages 1609-1618.

Nayef, N., and Ogier, J. M. (2015). Text zone classification using unsupervised feature learning. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, pages 776-780.

Neha, N. (2012). Language independent robust skew detection and correction technique for document images, pages 111-115.

Nezamabadi-pour, H., and Saryazdi, S. (2005). An Efficient Method for Document Image Enhancement, International Symposium on Telecommunications, pages 175-180.

Nor, D. M., Omar, R., Jenu, M. Z. M., and Ogier, J. M. (2011). Image segmentation and text extraction: application to the extraction of textual information in scene images. In International Seminar on Application of Science Mathematics (ISASM). pages 01-08.

Okun, O., Dœrmann, D., and Pietikainen, M. (1999). Page segmentation and zone classification: the state of the art. pages 01-34.

Oliveira, S. A., Seguin, B., and Kaplan, F. (2018). dhsegment: A generic deep-learning approach for document segmentation, 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 07–12.

Pal, U., and Chaudhuri, B. B. (1999). Automatic separation of machine-printed and hand-written text lines. In Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR, India, pages 645-648.

Pallipamu, V., Reddy, T. K., and Varma, S. P. (2014). A Survey on Digital Signatures, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 3, pages 7243-7246.

Parveen, R., Kulkarni, S., and Mytri, V. D. (2020). Automated extraction and discrimination of open land areas from IRS-1C LISS III imagery, International Journal of Computers and Applications, Vol 42, Issue 7, pages 676-685.

Patel, R. J., and Mitra, S. K. (2015). Extracting text from degraded document image, Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pages 01-04.

Pati, P. B., Raju, S. S., Pati N., and Ramakrishnan A. G. (2004). Gabor filters for document analysis in Indian bilingual documents. In International Conference on Intelligent Sensing and Information Processing, pages 123-126.

Peng, X., Setlur, S., Govindaraju, V., Sitaram, R., and Bhuvanagiri, K. (2009). Markov random field based text identification from annotated machine printed documents. In 2009 10th International Conference on Document Analysis and Recognition, IEEE, pages 431-435.

Perumal, S., and Velmurugan, T. (2018). Preprocessing by Contrast Enhancement Techniques for Medical Images, International Journal of Pure and Applied Mathematics, Vol. 118, No. 18, pages 3681-3688.

Pham, T. D. (2003). Unconstrained logo detection in document images, Pattern recognition, Vol. 36, Issue 12, pages 3023-3025.

Pham, T.A., Delalandre, M., and Barrat, S. (2011). A contour based method for logo detection. International Conference on Document Analysis and Recognition, pages 718-722.

Plamondon, R., and Srihari, S. N. (2000). Online and off-line handwriting recognition: a comprehensive survey. IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, Issue 1, pages 63-84.

Pramanik, R., and Bag, S. (2021). A novel skew correction methodology for handwritten words in multilingual multi-oriented documents, Multim. Tools Appl. 80, pages 27323–27342.

Purba, A. M., Harjoko, A., and Wibowo, M. E. (2019). Text Detection in Indonesian Identity Card Based On Maximally Stable Extremal Regions, IJCCS (Indonesian Journal of Computing and Cybernetics Systems), pages 177-188.

Puri, S., and Singh, S. P. (2016). Text recognition in bilingual machine printed image documents—Challenges and survey: A review on principal and crucial concerns of text extraction in bilingual printed images. In 10th International Conference on Intelligent Systems and Control (ISCO), IEEE, pages 01-08.

Puri, S., and Singh, S. P. (2020). A Fuzzy Matching based Image Classification System for Printed and Handwritten Text Documents, JITR Vol.13, No.2, pages 155-194.

Qi, X., Ma, L., Sun, C., and Liu, J., (2011). Fast skew angle detection algorithm for scanned document images. Third Pacific-Asia Conference on Circuits, Communications and System (PACCS), 2011, pages 01–04.

Radhika, K. S., and Gopika, S. (2015). Online and offline signature verification: a combined approach, Procedia Computer Science, 46, International Conference on Information and Communication Technologies (ICICT), pages 1593-1600.

Rajesh, T. M. (2015). An application of GMM algorithm in signature skew detection, https://www.researchgate.net/publication/325137210, pages 01-15.

Rani, A., and Singh, H. (2015). A review on various techniques for skew detection and correction in handwritten text documents. IJREAT, Volume 3, pages 58-66.

Ranota, H. K., and Kaur, P. (2014). Review and Analysis of Image Enhancement Techniques, International Journal of Information and Computation Technology. ISSN 0974-2239 Vol. 4, No. 6, pages 583-590.

Ravikumar, M., and Boraik, O. A. (2020). Low Pass Filter-Based Enhancement of Arabic Handwritten Document Images, ICTIS, Smart Innovation, Systems and Technologies, vol 195. Springer, Singapore. pages 01-07.

Ravikumar, M., Rachana, P. G., Shivaprasad, B. J., and Shivakumar, G. (2017). Segmentation of Words from unconstrained multilingual handwritten documents. J. Innov. Comput. Sci. Eng. 6, pages 26-27.

Ravikumar, M., Shivaprasad, B. J., and Guru, D. S. (2020). Enhancement of MRI Brain Images Using Fuzzy Logic Approach, Recent Trends in Image Processing and Pattern Recognition. RTIP2R, pages 01-07.

Redmon, J., and Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7263-7271.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, pages 1137-1149

Reshi, I. A. (2017). New Techniques Used for Image Enhancement, IOSR Journal of VLSI and Signal Processing (IOSR-JVSP), Vol. 7, Issue 6, Ver. I, pages 18-22.

Riba, P., Dutta, A., Goldmann, L., Fornés, A., Ramos, O., and Llados, J. (2019). Table detection in invoice documents by graph neural networks. In 2019 International Conference on Document Analysis and Recognition, pages 122-127.

Romit, B., Goyal, N., Debapriya, G., and Farhana, J. Z. (2009). Signature Authentication, www.researchgate.net/publication/324970502, pages 01-11.

Rosli, M. M., Tempero, E., and Luxton-Reilly, A., (2016). What is in our datasets? Describing a structure of datasets, ACSW Multiconference, Publication rights licensed to ACM, pages 01-10.

Sabourin, R., and Plamondon, R. (1988). Segmentation of handwritten signature images using the statistics of directional data. In 9th International Conference on Pattern Recognition, IEEE Computer Society, pages 282-285.

Saha, R., Mondal, A., and Jawahar, C. V. (2019). Graphical object detection in document images. In 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, pages 51-58.

Sakila, A., and Vijayarani, D. S. (2017). Skew detection and correction in the document image, International Journal of Innovative Research in Science Engineering and Technology, pages 17457- 17465.

Salagar, R. D., and Patil, P. B. (2020). Application of rlsa for skew detection and correction in kannada text images', Fourth International Conference on Computing Methodologies and Communication (ICCMC), pages 785–788.

Sarath, K., and Sreejith, S. (2017). Image Enhancement Using Fuzzy Logic, IOSR Journal of Electronics and Communication Engineering, pages 34-44.

Sarker, S., Chowdhury, S., Laha, S., and Dey, D. (2012). Use of non-local means filter to denoise image corrupted by salt and pepper noise. Signal and Image Processing, Vol. 3, Issue 2, pages 223-235.

Sasirekha, D., and Chandra, E. (2012). Enhanced techniques for PDF image segmentation and text extraction. arXiv preprint arXiv:1210.0347, pages 01-05.

Sattar, F., and Tay, D. B. (1999). Enhancement of document images using multiresolution and fuzzy logic techniques, IEEE Signal Processing Letters, Vol. 6, Issue 10, pages 249-252.

Saxena, N., and Parveen, H. (2019). Text extraction systems for printed images: a review. International Journal of Advanced Studies of Scientific Research, Vol. 4, Issue 2, pages 513-519.

Schreiber, S., Agne, S., Wolf, I., Dengel, A., and Ahmed, S. (2017). Deepdesrt: Deep learning for detection and structure recognition of tables in document images. International conference on document analysis and recognition, Vol. 1, pages 1162-1167.

Shafait, F., and Smith, R. (2010). Table detection in heterogeneous documents. In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, pages 65-72.

Shah, L., Patel, R., Patel, S., and Maniar, J. (2014). Skew Detection and Correction for Gujarati Printed and Handwritten Character using Linear Regression, International Journal, Vol. 4, Issue 1, pages 642-648.

Shakunthala, B. S., and Naveen, N. B. S. (2017). Unconstrained handwritten kannada documents leading to line and word segmentation, International journal of engineering research and technology, Vol. 5, Issue 20, pages 01-05.

Shakunthala, B., and Pillai, C. (2021). Enhanced text line segmentation and skew estimation for handwritten kannada document, Journal of Theoretical and Applied Information Technology, Vol. 99, Issue 1, pages 196–206.

Sharma, N., Mandal, R., Sharma, R., Pal, U., and Blumenstein, M. (2018). Signature and logo detection using deep CNN for document image retrieval. In 16th International Conference on Frontiers in Handwriting Recognition, pages 416-422.

Sheetala, S. C., and Dixit, U. D. (2015). Detection and Logo Based Document Image Retrieval: A Review, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue 5, pages 2822-2827.

Shekar, B. H., and Smitha, M. L. (2015). Text Localization in Video/Scene Images using Kirsch Directional Masks, IEEE(ICACCI), pages 1436-1440.

Shetty, S., Srinivasan, H., Beal, M., and Srihari, S. (2007). Segmentation and labeling of documents using conditional random fields, In Document Recognition and Retrieval XIV, Vol. 6500, SPIE, pages 256-264.

Shi, Z., and Govindaraju, V. (2004). Fuzzy Run length in Document Image Processing, Center of Excellence for Document Analysis and Recognition (CEDAR), Elsevier Science, pages 01-13.

Shi, Z., Setlur, S., and Govindaraju, V. (2004). Digital enhancement of palm leaf manuscript images using normalization techniques. In 5th International Conference On Knowledge Based Computer Systems, pages 19-22.

Shi, Z., Setlur, S., and Govindaraju, V. (2011). Image Enhancement for Degraded Binary Document Images, International Conference on Document Analysis and Recognition, pages 895-899.

Shirazi, A. A., Dehghani, A., Farsi, H., and Yazdi, M. (2017). Persian Logo Recognition using Local Binary Patterns. 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA), pages 258-261.

Shirdhonkar, M. S., and Kokare, M. B. (2010). Discrimination between printed and handwritten text in documents. IJCA Special Issue on. Recent Trends in Image Processing and Pattern Recognition, pages 131-134.

Shivakumara, P., Kumar, G. H., Guru, D. S., and Nagabhushan, P. (2005). A novel technique for estimation of skew in binary text document images based on linear regression analysis, Sadhana, Vol. 30, Issue 1, pages 69-85.

ShobhaRani, N., and Vasudev, T. (2015). 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, Vol. 6, Issue 2, pages 458-465.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, pages 01-14.

Singh, A.K., and Singh, P. (2014). Text Extraction from Live Captured Image with Diversified Background using Edge Based and K-Means Clustering. International Journal of Innovations in Engineering and Technology (IJIET), Vol. 3, Issue 11, pages 11-17.

Singh, C., Bhatia, N., and Kaur, A. (2008). Hough transform based fast skew detection and accurate skew correction methods. Pattern Recognition., 41, pages 3528-3546.

Singha, C., Bhatiab, N., and Amandeep Kaur, C. (2008). Hough transform based fast skew detection and accurate skew correction methods. ELSEVIER-Pattern Recogn. 41, pages 3528–3546.

SmitaPatil, A., and Mohite-Patil, T. B. (2016). Logo Detection and Recognition Using Context Dependency for Image, International Research Journal of Engineering and Technology (IRJET), pages 90-93.

Song, Y., Chen, J., Xie, H., Chen, Z., Gao, X and Chen, X. (2017). Robust and parallel Uyghur text localization in complex background images, Machine Vision and Applications, Vol. 28, Issue 7, pages 755-769.

Soora, N. R., and Deshpande, P. S. (2018). A novel local skew correction and segmentation approach for printed multilingual Indian documents, Alexandria engineering journal, Vol. 57, Issue 3, pages 1609-1618.

Soumya, K. R., and Chacko, A. (2014). Text Extraction from Images: A Survey., International Journal of Advances in Computer Science and Technology, Vol. 3, No. 2, pages 100-104.

Srivastva, R., Raj, A., Patnaik, T., and Kumar, B. (2013). A survey on techniques of separation of machine printed text and handwritten text, Vol. 2 Issue 3, pages 552-555.

Sugapriya, C., (2017). Quality Improvement of Image Processing Using Fuzzy Logic System, Advances in Computational Sciences and technology, Vol. 10, Issue 6, pages 1849-1855.

Sun, C., and Si, D. (1997). Skew and slant correction for document images using gradient direction. Fourth International Conference on Document Analysis and Recognition, pages 142-146.

Sun, L., Huo, Q., Jia, W., and Chen, K. (2015). A robust approach for text detection from natural scene images. Pattern Recognition, Vol. 48, Issue 9, pages 2906-2920.

Tan, L., and Jiang, J. (2019). Adaptive Filters and Applications, Digital Signal Processing. doi:10.1016/b978-0-12-815071-9.00009-9, pages 421-474.

Tang, Y. Y., Lee, S. W., and Suen. C. Y. (1996). Automatic Document Processing: A Survey. Pattern recognition, Vol. 29, pages 1931-1952.

Tarek, M., Hamza, T., and Radwan, E. (2010). Off-Line Handwritten Signature Recognition Using Wavelet Neural Network. International Journal of Computer Science and Information Security, Vol. 8, Issue 6, pages 01-09.

Thakur, A., and Mishra. D. (2015). Fuzzy contrast mapping for image enhancement. 2nd International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, pages 549-552.

Thangaraj, M., and Sivakami, M. (2018). Text classification techniques: a literature review. Interdisciplinary Journal of Information, Knowledge, and Management, 13, 117. pages 117-135.

Tombre, K., Tabbone, S., Pélissier, L., Lamiroy, B., and Dosch, P. (2002). Text/graphics separation revisited. In International Workshop on Document Analysis Systems. Springer, Berlin, Heidelberg. pages 200-211.

Tran, T. A., Na, I. S., and Kim, S. H. (2015). Separation of text and non-text in document layout analysis using a recursive filter. KSII Transactions on Internet and Information Systems (TIIS), Vol. 9, Issue 10, pages 4072-4091.

Tupaj, S., Shi, Z., Chang, C. H., and Alam, H. (1996). Extracting tabular information from text files. EECS Department, Tufts University, Medford, USA, pages 01-18.

Umer, S., Mondal, R., Pandey, H. M., and Rout, R. K. (2021). Deep features based convolutional neural network model for text and non-text region segmentation from document images, Applied Soft Computing, Vol. 113, pages 01-09.

Vaquar, M., Handa, P., and Rawat, S. (2019). A comparative analysis of image enhancement techniques. In International Conference on Advances in Engineering Science Management and Technology (ICAESMT), Uttaranchal University, Dehradun, India, pages 01-06.

Viana, M. P., and Oliveira, D. A. (2017). Fast CNN-based document layout analysis. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 1173-1180.

Viana, M. P., and Oliveira, D. A. B. (2017). Fast CNN-based document layout analysis. IEEE International Conference on Computer Vision Workshops (ICCVW), pages 1173–1180.

Wagdy, M., Faye, I., and Rohaya, D. (2014). Document image skew detection and correction Method based on extreme points, Centre of intelligent signal and imaging research (IEEE), Vol. 6, Issue 1, pages 01-05.

Wang, H. (2010). Document logo detection and recognition using bayesian model, 20th International Conference on Pattern Recognition, IEEE, pages 1961-1964.

Wang, H., and Chen, Y. (2009). Logo detection in document images based on boundary extension of feature rectangles, In 2009 10th International Conference on Document Analysis and Recognition, IEEE, pages 1335-1339.

Wang, H., Pan, C., Guo, X., Ji, C., and Deng, K. (2021). From object detection to text detection and recognition: A brief evolution history of optical character recognition, Wiley Interdisciplinary Reviews: Computational Statistics 13, pages 01–32.

Watts, N., and Rani, J. (2014). Performance evaluation of improved skew detection and correction using FFT and median filtering, International Journal of Computer Applications, pages 07–16.

Winiarti, S., Ismi, D. P., and Prahara, A. (2017). Image enhancement using piecewise linear contrast stretch methods based on unsharp masking algorithms for leather image processing, 3rd International Conference on Science in Information Technology (ICSITech), pages 669-673.

Wu, H., Zou, B., Zhao, Y. Q., Chen, Z., Zhu, C., and Guo, J. (2016). Natural scene text detection by multi-scale adaptive color clustering and non-text filtering, Neuro computing, 214, pages 1011-1025.

Xiong, J., Yu, D., Wang, Q., Shu, L., Cen, J., Liang, Q., and Sun, B. (2021). Application of Histogram Equalization for Image Enhancement in Corrosion Areas, Shock and Vibration, Vol. 2021, pages 01-13.

Yadav, V., and Ragot, N. (2016). Text extraction in document images: highlight on using corner points, In 12th IAPR Workshop on Document Analysis Systems (DAS), IEEE, pages 281-286.

Yahya, S. R., Abdullah, S. S., Omar, K., Zakaria, M. S., and Liong, C. Y. (2009). Review on image enhancement methods of old manuscript with the damaged background, In 2009 International Conference on Electrical Engineering and Informatics, Vol. 1, IEEE, pages 62-67.

Yi, X., Gao, L., Liao, Y., Zhang, X., Liu, R., and Jiang, Z. (2017). CNN based page object detection in document images. 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Vol. 1, IEEE, pages 230-235.

Zagoris, K., Pratikakis, I., Antonacopoulos, A., Gatos, B., and Papamarkos, N. (2014). Distinction between handwritten and machine-printed text based on the bag of visual words model. Pattern Recognition, Vol. 47, Issue 3, pages 1051-1062.

Zeiler, M. D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision. Springer, Cham. pages 818-833.

Zhang, X. W., Zheng, X. B., and Weng, Z. J. (2008). Text extraction algorithm under background image using wavelet transforms, In 2008 International Conference on Wavelet Analysis and Pattern Recognition, Vol. 1, pages 200-204.

Zhang, Y., and Tan, K. K. (2009). Text extraction from images captured via mobile and digital devices. International Journal of Computational Vision and Robotics, Vol. 1, Issue 1, pages 34-58.

Zhang, Y., Zhu, M., Wang, D., and Feng, S. (2014). Logo detection and recognition based on classification, In International Conference on Web-Age Information Management Springer, Cham, pages 805-816.

Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., and Zhang, D. (2019). Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging, Sensors (Basel, Switzerland) 19, pages 01–13.

Zhao, Z. Q., Zheng, P., Xu, S. T., and Wu, X. (2019). Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, Vol. 30, Issue 11, pages 3212-3232.

Zheng, Y., Li, H., and Doermann, D. (2002). The segmentation and identification of handwriting in noisy document images, In International Workshop on Document Analysis Systems, Springer, Berlin, Heidelberg, pages 95-105.

Zhou, F., Jia, Z., Yang, J., and Kasabov, N. (2017). Method of improved fuzzy contrast combined adaptive threshold in NSCT for medical image enhancement, BioMed Research International, Vol. 2, pages 01-10.

Zhu, G., Zheng, Y., Doermann, D., and Jaeger, S. (2008). Signature detection and matching for document image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, Issue 11, pages 2015-2031.