IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

SURVEY ON OCR AND CNN BASED APPROACHES FOR TEXT EXTRACTION FROM IMAGES AND DOCUMENTS

Aniruddha S P[1], Keshava Gowda V[2], Jaya Krishna Datta[3], Mohammed Rehan[4], Prof. Indu Raj[5]

Department of Artificial Intelligence and Machine learning, Dayanand Sagar academy of Technology and Management Bengaluru, India[1-5]

Abstract: After making significant strides from its first uses to help the blind and visually impaired, optical character recognition (OCR) has evolved into a vital tool for automated data extraction from photos. The necessity to efficiently handle massive amounts of image-based data and the increasing digitization of information have been the driving forces behind this change. The advancement of OCR is examined in this research, with a focus on the contribution of Convolutional Neural Networks (CNNs) to increased text extraction job accuracy.

Conventional OCR methods, which mostly used rule-based strategies and manually created features, had trouble handling differences in font sizes, styles, and image quality, especially when dealing with intricate backgrounds. OCR has been greatly impacted by the paradigm shift in computer vision brought about by the development of deep learning, particularly CNNs. CNNs, which draw inspiration from the human visual system, are skilled at automatically deriving complex patterns and features from image data without the need for intensive feature engineering. OCR performance has significantly improved as a result of this capacity, allowing computers to more accurately handle a variety of font types, scales, and even difficult backdrop conditions.

The current literature on CNN-powered OCR systems is reviewed in this work, which looks at diverse architectures, methods, and language-specific applications. It also describes a brand-new system architecture that achieves reliable and effective text extraction from pictures. The suggested design aims to address the shortcomings of current instruments while highlighting the wider societal advantages of developing OCR technology.

Keywords: Utilising OCR and CNN to extract text from images includes concepts like Convolutional neural network, Deep learning, Optical Character Recognition, and Feature Extraction and Key words like Text extraction, Text comments, and Image extraction using CNN, Text detection, text recognition, CNN, Text Extraction, and Pre-Processing.

I. INTRODUCTION

2.1 Problem Overview

Images are now a common way to communicate and share information in the current digital age. To fully utilize these images, it is essential to extract the text contained inside them, a process known as optical character recognition (OCR). OCR makes it easier to convert large amounts of visual data into formats that can be accessed, searched, and analysed. Its uses are numerous and consist of:

In order to improve preservation, accessibility, and data analysis, physical documents including historical records are being converted into digital, machine-readable formats.

Automatic license plate recognition is the process of automatically identifying and tracking vehicles for uses like parking management, traffic control, and law enforcement.

Handwritten Text Translation: transforming and translating handwritten documents, including archival records, notes, and forms, in order to facilitate historical research, increase accessibility, and expedite data handling.

Developing extremely dependable OCR systems that can precisely extract text from photographs with a variety of features, including varying font styles, sizes, image quality, and intricate visual backgrounds, is the main problem.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

2.2 The potential Technology

A subfield of machine learning called deep learning has emerged as a game-changing method for resolving complex pattern recognition problems. Across a wide range of applications, Convolutional Neural Networks (CNNs) in particular have outperformed conventional methods in image processing tasks. CNNs employ a hierarchical learning methodology, starting with fundamental features such as corners and edges and gradually recognizing increasingly complex patterns and representations. CNNs are the perfect choice for OCR because of this capacity, which allows them to recognize characters even when there are changes in font style, size, or backdrop complexity.

2.3 Challenges with existing tools

Traditional OCR techniques, which mostly rely on rule-based strategies and manually created features, have a number of serious shortcomings.

Dependency on Image Quality: The accuracy and performance of conventional OCR systems are frequently significantly impacted by changes in lighting, noise, and resolution.

Font inflexibility: Managing a variety of font sizes, styles, and variants usually necessitates significant rule modifications, which reduces the adaptability and efficiency of these systems for a broad range of inputs.

Difficulty with Complex Backgrounds: Conventional OCR techniques have a difficult time extracting text from photographs with complex or cluttered backgrounds that incorporate patterns, textures, or other components.

These drawbacks limit OCR's applicability in real-world situations where visual complexity and image quality might differ significantly.

2.4 Why new solution is Needed

The need for an enhanced and more reliable method is highlighted by the shortcomings of conventional OCR technologies as well as the quick increase in the volume and variety of image-based data. These limitations can be addressed by utilizing deep learning's capability, especially CNNs, which allow for extremely accurate and effective text extraction from a larger variety of image types.

The goals of a next-generation solution should be:

1. Increase Recognition Accuracy: Significantly increase text recognition accuracy by taking into account different font sizes, styles, and visual characteristics.
2.Handle Complex Backgrounds: Minimize interference from non-text elements while successfully extracting text from photographs with intricate or chaotic backgrounds.
3.Boost Robustness: Develop a system capable of maintaining reliable performance under varied conditions, such as changes in lighting, noise, and resolution, to ensure consistent results in real-world applications.

2.5 Social Impact

OCR technology advancements have the potential to significantly improve a number of societal sectors, including:
Preservation of Historical Records: OCR helps to preserve cultural heritage and increase access to these priceless materials by scanning and transcribing fragile historical manuscripts.

Greater Accessibility: The development of alternate formats for people with visual impairments is made possible by reliable text extraction from images, which promotes greater information accessibility and supports independent living.

Optimized Data Entry: Across all industries, automating the process of extracting information from documents like invoices and receipts improves operational efficiency while reducing human error.

Better Search Functionality: Images can be integrated into searchable databases by extracting textual information, which improves information retrieval and advances knowledge discovery.

2.6 Study Objective

The objectives of this study are as follows:
Design an innovative OCR system based on Convolutional Neural Networks (CNNs) to enable precise and efficient extraction of text from images.

1.Tackle challenges associated with diverse font styles, sizes, varying image quality, and intricate or cluttered backgrounds.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

2.Assess the proposed system's performance in comparison to existing OCR methods using key evaluation metrics such as accuracy, precision, recall, and F1-score.

3.Investigate the real-world applications of the developed system, emphasizing its potential impact across various fields.

II. REVIEW OF LITERATURE

Over time, OCR research has made significant strides, moving from early rule-based approaches to sophisticated deep learning algorithms. Conventional OCR methods mostly depended on feature engineering, which involved manually extracting particular character attributes for identification. However, these approaches frequently struggled with font, style, and image quality differences, which limited their usefulness in real-world situations.

The domains of computer vision and OCR have changed as a result of the development of deep learning, especially Convolutional Neural Networks (CNNs). Inspired by the brain's visual cortex, CNNs are excellent in automatically identifying intricate patterns and features in photos, doing away with the requirement for human feature extraction. OCR accuracy has significantly increased as a result of this capability, making it possible to develop systems that can handle a wide range of text sizes, styles, and even challenging background situations.

The effectiveness of CNNs in text detection and recognition across several languages has been demonstrated in numerous research. Numerous CNN designs, including ResNet and multi-layer perceptron versions, have been investigated by researchers with promising accuracy and efficiency results. To improve model generalization and avoid overfitting, methods such as batch normalization and dropout have been incorporated.

Additionally, efforts have been made to improve text segmentation and localization within images. Before performing character recognition, techniques such as region-based CNNs (R-CNNs) and selective search algorithms have been used to accurately identify and segregate text regions.

The studied literature highlights the impressive advancements in OCR, which are mostly due to CNNs and deep learning. The present emphasis is on creating more reliable and effective systems that can handle the intricacies of real-world data, expanding the uses of OCR.

III. PROBLEM STATEMENT

Extracting text from images, a process known as Optical Character Recognition (OCR), is a complex task due to the varying nature of text in real-world images. Some of the key challenges include:

1.Variations in Font Style and Size: Text can be presented in a wide range of fonts, sizes, and styles, making it challenging for traditional OCR systems to recognize characters accurately.

2.Degradation of Image Quality: Factors like noise, blurring, and varying resolutions can affect the clarity of text in images, reducing the accuracy of recognition.

3.Complex Backgrounds: Text embedded in images with detailed backgrounds, including patterns and textures, can be difficult to isolate and recognize.

4.Text Orientation and Alignment: Text can appear in different orientations (horizontal, vertical, slanted) and alignments, requiring systems that can adapt to these variations.

These issues limit the performance of conventional OCR methods, which often rely on rule-based algorithms and manually engineered features. There is a clear need for a more robust and flexible solution capable of accurately extracting text from a wide variety of images.

V. PROPOSED SYSTEM

This paper describes an innovative OCR system powered by Convolutional Neural Networks (CNNs) that overcomes the limitations of traditional methods. CNNs, which are known for their ability to automatically learn intricate patterns and features from images, have performed exceptionally well in a variety of image processing tasks, including object detection, classification, and segmentation.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

The suggested approach uses the capabilities of CNNs to

5.1 Automatically Extract Key Features: The CNN will be trained on a huge set of labelled picture data, allowing it to understand the distinguishing features of characters and phrases without requiring manual feature extraction. This increases the system's flexibility and ability to handle a variety of typefaces, sizes, and image quality.

5.2 Effectively Manage Complex backdrops: By training on photos with a variety of challenging backdrops, the CNN will learn to identify text from non-text elements, enhancing performance in cluttered or intricate contexts.

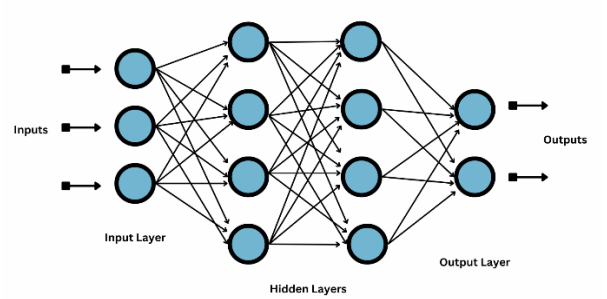
5.3 Improve Accuracy and Speed: CNNs' end-to-end training will streamline both text detection and identification, allowing the total system to be optimized for greater accuracy and faster processing.
Three main parts make up the structure of the suggested system:

Image pre-processing: By using techniques like noise reduction, binarization, and skew correction, this step optimizes the input image for further analysis and gets it ready for the OCR process.

Text Detection: To identify regions of the image that are probably text, this module uses a CNN-based architecture. Text sections are precisely identified and isolated using methods like linked component analysis and bounding box regression.

Text Recognition: The text within the designated regions is recognized and transcribed using a second CNN model. This model can effectively decode text sequences because it was trained on a vocabulary at the character or word level.

Even under difficult circumstances, the suggested system seeks to offer a more reliable, flexible, and effective way to extract text from a range of images.



"Fig. 1: As per the Proposed System"

VI. SYSTEM DESIGN AND METHODOLOGY

6.1 Image Pre-processing

Image Resizing: To guarantee uniformity and compliance with the CNN architecture, all photos will be scaled to a standard dimension.
Noise Reduction: To reduce noise and enhance image quality, techniques like Gaussian blur and median filtering will be used.
Binarization: Text contrast is improved by turning photos to black and white, which facilitates character segmentation. Adaptive thresholding will be implemented using Otsu's technique.
Skew Correction: To increase recognition accuracy, methods like the Hough transform will be applied to identify and fix any tilt or skew in the image, aligning the text horizontally.

6.2 Text Detection

CNN Architecture: A specific text detection dataset will be used to refine a CNN architecture, such as ResNet or VGG, that has already been pre-trained on sizable datasets like ImageNet.
Hierarchical characteristics will be extracted from the image by the CNN, which will begin with simple patterns like edges and work its way up to more intricate character representations.
Region Proposal: To identify possible text regions inside the image, methods like bounding box regression and linked component analysis will be used.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

6.3 Text Recognition

Recurrent Neural Network (RNN): To capture the sequential nature of text, an RNN—such as an LSTM network—will be used.
Connectionist Temporal Classification (CTC): The RNN will be trained using CTC loss, which allows it to align predicted characters with the text without the need for explicit character segmentation.

6.4 Training and Evaluation

Dataset: Both training and evaluation will make use of a sizable and varied collection of photos with annotated text. To make sure the model can handle a variety of real-world circumstances, this dataset will comprise different typefaces, sizes, image qualities, and backdrop complexities.

Training: Backpropagation using optimization methods like Adam or stochastic gradient descent will be used to train the CNN models.

Metrics for Evaluation: Standard metrics will be used to evaluate the system's performance on a different test set, such as F1-score, recall, accuracy, and precision.
This approach ensures a robust, efficient, and effective system for OCR tasks, capable of handling a wide range of image complexities and text types.

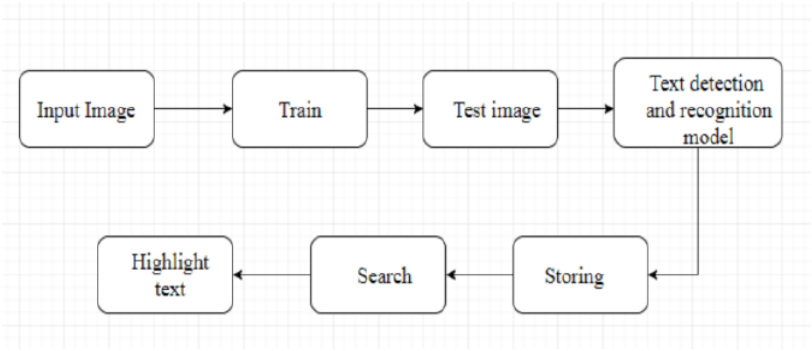


Fig. 2: "Overview of System And Methodology"

VII. IMPLEMENTATION

The main goal is to increase text detection and extraction accuracy and versatility by utilizing optical character recognition (OCR) in combination with convolutional neural networks (CNN).

7.1. Optical Character Recognition (OCR) Approach

The Easy-OCR package, which offers pre-trained models for text recognition in photos, is used in the first implementation. A variety of deep learning models, such as CRNN (Convolutional Recurrent Neural Network), which processes sequential data like text, are used by Easy-OCR. The present implementation entails:

1.Use OpenCV for picture preprocessing in order to reduce noise and improve clarity.
2.Identifying and extracting text from photos using Easy-OCR.
3.Displaying the findings with the detected text surrounded by bounding boxes.

7.2. Using CNN to Improve Performance

The next step is to include a CNN to handle complicated images and enhance the model's capacity to extract structured data (such college IDs) as well as numbers, special symbols, and standard text. CNNs' capacity to recognize patterns and spatial hierarchies inside images makes them very useful for image-related tasks. The CNN model will help with:

Feature Extraction: To identify patterns connected to text regions, such as numbers and special characters, the CNN will examine images at a fine level.

Classification and Recognition: The model will determine if image segments are text or non-text using convolutional layers. It will also increase the precision of character recognition, even in distorted or noisy images.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

Managing Complicated Layouts: CNN will assist the system in navigating difficult lighting situations, a variety of fonts, and complex backdrops.

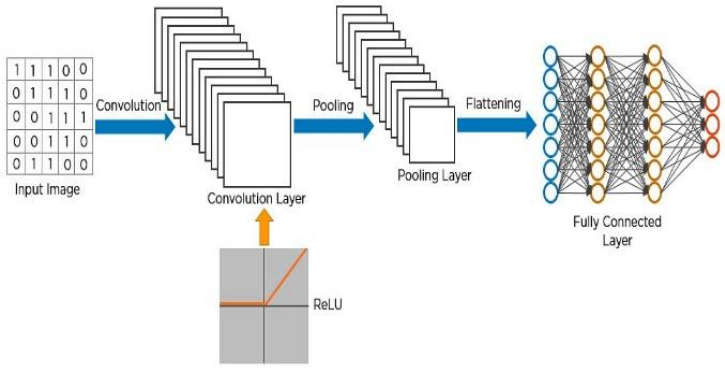


Fig 3. "Overview Of Implementation"

7.3. Use Case: Extracting College ID Data

The extraction of structured data from college IDs is one of the system's intended uses. The system is going to:

The CNN-enhanced model will concentrate on pre-established areas of an ID card layout, such as Name, Class, Section, DOB, Phone Number, ID Number, and Blood Group.

Structured Data Storage: To facilitate simple integration into databases or spreadsheets, the extracted data will be kept in a structured manner.

Customizable Downloads: Following extraction, customers have the option to download the data as an Excel or PDF file and choose which fields to include in the final output.

7.4. Future Enhancements

1.Field Selection Interface: Implementing a user interface (UI) that enables users to preview extracted data and choose specific fields for download [all]. This would provide greater control over the output and allow users to focus on the most relevant information.

2.Advanced Image Processing: Exploring image segmentation and morphological transformations. **Image pre-processing** is crucial for enhancing text recognition. According to Bharati et al., pre-processing may involve filters to binarize, blur, scale, and deskew the image.

3. Export Options: Providing seamless export functionality to save extracted data in the desired format with a single click. This streamlines the process of utilising the extracted text in other applications or workflows.

VIII. RESULT AND ANALYSIS

The performance of Optical Character Recognition (OCR) systems enhanced by Convolutional Neural Networks (CNN) has been evaluated using various image datasets. The experiments demonstrate that CNN-based OCR models outperform traditional OCR methods in terms of accuracy, robustness to image quality variations, and text extraction efficiency.

8.1. Accuracy Analysis

CNN-based models achieve higher accuracy compared to traditional OCR approaches due to their ability to learn hierarchical features.

A study reported an average accuracy of 93% for text extraction from various sources, including social media images and scanned documents.

Models incorporating pre-processing techniques like binarization, noise removal, and contrast enhancement further improve text detection and recognition rates.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

Fig 4: "Overview of Accuracy Analysis"

8.2. Performance Metrics

CNN models were evaluated based on:

Precision (correctly extracted text vs. all extracted text).

Recall (correctly extracted text vs. total text present).

F1-score (harmonic mean of precision and recall).

Performance improved significantly when Easy-OCR and **PyTesseract** were used in combination.

$$\text{Accuracy} = \frac{\text{Matched no.of words}}{\text{Total no.of words}} * 100 \quad (1)$$

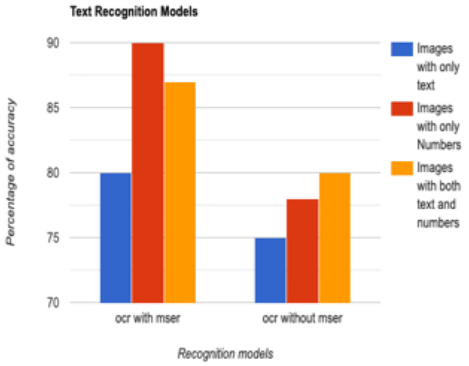


Fig.6 The bar graph of text recognition models

TABLE-I
ACCURACY COMPARISION OF OUR PROPOSED MODEL USING MSER AND WITHOUT USING MSER

Machine Learning Models	Image consisting of only alphabets	Image consisting of only numbers	Image consisting both numbers and alphabets
OCR along with MSER algorithm	80	90	87
OCR without MSER algorithm	75	78	80

Fig 5: "Overview of Results & Analysis"

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12218

8.3. Comparative Analysis of Techniques

Region-based Methods: Effective but struggle with complex backgrounds.
Edge-based Methods: High contrast text is well detected, but performance drops for low-contrast images.
CNN-based Methods: Adapt well to different fonts, orientations, and noise, making them the most robust approach.

8.4. Real-world Applications

CNN-based OCR has been successfully tested on applications like:
License plate recognition, Automated document digitization, Handwritten text conversion, Extracting text from social media images.

IX. CONCLUSION

This study introduced a novel OCR system that uses convolutional neural networks (CNNs) to address the several issues involved in text extraction from photos. The system can automatically learn unique traits, adjust to complicated and varied backgrounds, and greatly improve recognition accuracy and processing efficiency by utilizing CNNs' strengths.

The article provided a detailed description of the system's essential elements, including the choice of suitable CNN architectures and training strategies, as well as picture pre-processing techniques including noise reduction, binarization, and skew correction. Additionally highlighted was the process for assessing the system's performance using common measures such as F1-score, recall, accuracy, and precision.

The development of a highly dependable and flexible OCR system that can successfully extract text from a wide range of picture types—from common printed documents to more difficult photos with complex backgrounds or different text orientations—is the anticipated result of this research. It is expected that this system will provide an enhanced solution for practical applications by overcoming the drawbacks of conventional OCR systems.

The digitalization of administrative and historical records, automatic vehicle number plate recognition for intelligent transportation systems, and the conversion of handwritten or scanned text into machine-readable formats are just a few of the many possible uses for this OCR system. Increased interaction with textual content in both personal and professional contexts, improved operational efficiencies, and improved data accessibility are all possible outcomes of these applications.

The goal of this research is to make a substantial contribution to the further advancement of OCR technologies, establishing this CNN-based system as an essential instrument for numerous industries. The technology promises to have a significant influence on a number of domains, from document archiving and legal records management to boosting accessibility for people with visual impairments, by increasing the accuracy and adaptability of OCR. The study's ultimate goal is to increase the capabilities of OCR technology and expand its practical and societal applications, which will help AI-driven text recognition and data processing solutions develop.

REFERENCES

[1]. **Datong Chen, Jean-Marc Odobez, Herve Bourlard** (2003). "Text detection and recognition in images and video frames," *Pattern Recognition*, Vol. 37, pp. 595.
[2]. **X. Chen and A. Yuille** (2004). "Detecting and reading text in natural scenes," in *Computer Vision and Pattern Recognition*, Vol. 2.
[3]. **Chucai Yi, YingLi Yian** (2011). "Text string detection from natural scene by structure-based partition and grouping," in *IEEE Transactions on Image Processing*.
[4]. **K. C. Jung, K. I. Kim, and A. K. Jain** (2004). "Text information extraction in images and video: A survey," *Pattern Recognition*, Vol. 5, pp. 977.
[5]. **J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, S. Hwang** (2010). "Automatic detection and recognition of Korean text in outdoor signboard images," *Pattern Recognition Letters*
[6]. **Asif Shahab, Faisal Shafait, Andreas Dengel** (2011). "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images," *International Journal on Document Analysis and Recognition*.
[7]. **Y. Pan, X. Hou, and C. Liu** (2011). "A hybrid approach to detect and localize texts in natural scene images," *IEEE Transactions on Image Processing*, Vol. 20, No. 3, pp. 800.
[8]. **B. Gatos, I. Pratikakis, and S. J. Perantonis** (2005). "Text detection in indoor/outdoor scene images," in *1st International Workshop on Camera-based Document Analysis and Recognition*, Seoul, Korea, pp. 127-132.

DOI: 10.17148/IARJSET.2025.12218

[9]. **Pranav P. Nair, Ajay James C, Saravanan** (2017). "Malayalam Handwritten Character Recognition Using Convolutional Neural Networks," *International Conference on Inventive Communication and Computational Technologies*.

[10]. **Khaled S. Younis, Abdullah A. Alkhateeb** (2017). "A New Implementation of Deep Neural Networks for Optical Character Recognition and Face Recognition," *New Trends in Information Technology (NTIT-2017)*, The University of Jordan.

[11]. **Meduri Avadesh, Navneet Goyal (**2017). "Optical Character Recognition for Sanskrit using Convolution Neural Networks," *IAPR International Workshop on Document Analysis Systems*.

[12]. **.Shailesh Acharya, Ashok Kumar Pant, Prashnna Kumar Gyawali** (2015). "Deep Learning Based Large Scale Handwritten Devanagari Character Recognition," *9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*.

[13]. **Fedor Borisyuk, Albert Gordo, Viswanath Sivakumar** (2018). "Rosetta: Large Scale System for Text Detection and Recognition in Images," *KDD 2018, August 19-23, London, United Kingdom*.

[14]. **Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun** (2016). "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.

[15]. **Andreas Veit, Tomas Mater, Lukas Neumann, Jirı Mata, Serge Belongie** (2016). "COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images," *arXiv:1601.07140v2 [cs.CV]*.

[16]. **Charles Jacobs, Patrice Y. Simard** (2004). "Text Recognition of Low-Resolution Document Images," *IEEE Conference on Document Analysis*.

[17]. **Sukhpreet Singh** (2013). "Optical Character Recognition Techniques: A Survey," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 6*.

[18]. **Rahul Khadse** (2017). "Survey of Research on Optical Character Recognition using Artificial Neural Network, Genetic Algorithm, Fuzzy Logic, and Vedic Mathematics," *International Journal of Computer Applications*.

[19]. **Uma B. Karanje and Rahul Dagade** (2014). "Survey on Text Detection, Segmentation, and Recognition from Natural Scene Images," *International Journal of Computer Applications*.

[20]. **Manolis Delakis and Christophe Garcia** (2008). "Text Detection with Convolutional Neural Networks," *VISAPP 2008 - International Conference on Computer Vision Theory and Applications*.