IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SCENE IMAGE TO TEXT RECOGNITION IN MALAYALAM APP

Pioneering the Future of Talent Acquisition with AI-Powered Solutions

[1]Anaswara.C, [2]C.Swetha, [3]Smita unnikrishnan
[1]Student, [2]Student,[3]Assistant Professor (CSE)
[1]Computer Science and Engineering Department,
[1]Nehru College of Engineering and Research Centre (NCERC), Thrissur, India

Abstract: This research outlines an incredibly easy-to-use and effective technique for identifying Malayalam text in color, natural scene photos captured offline using a mobile phone camera. Important phases in text understanding from natural scene photographs are Malayalam text detection, text segmentation, skew correction of the discovered text, and character recognition.For a variety of applications, including text translation in other nations and assistive technology for the blind, text understanding in natural scene photos is crucial.Malayalam's high degree of complexity in comparison to other languages has made it difficult to learn. The experimental findings demonstrate that our approach can effectively extract and recognize text with little complexity, making it suitable for usage in mobile devices with constrained capabilities. Keywords: text detection, text segmentation, text recognition, and skew angle estimation.

Keywords: Skew angle estimation, text detection, text segmentation, text recognition

I. INTRODUCTION

Within computer vision, text recognition is a rapidly expanding field. Scene text recognition in deep learning has been developed based on the ongoing advancements in deep learning domains like computer vision, machine learning, and pattern recognition. Based on recognition algorithms, there are two branches of text recognition: segmentation-based recognition algorithms and recognition algorithms without segmentation requirements. The segmentation-based natural scene text recognition algorithm usually needs to locate the location of each character contained in the input text image, identify each character through a single character recognizer, and then combine all the characters into a string sequence to obtain the final recognition result. It extends the output layer of an RNN by: converting the data dependency of “segmenting” and label mapping relationships to extracting features according to a sliding time window, transforming the input– output relationship from one-to-one to many-to-one; adding blank characters, and performing deduplication and blank character removal operations on consecutive identical characters in the sequence output; reducing complexity and increasing speed by drawing on the forward and backward algorithms of the hidden Markov model (HMM) to compute the loss function; and using dynamic programming to compute the training paths, avoiding impractical exhaustive methods or violent.

II. LITERATURE SURVEY

A literature review is an overview of the works that recognized academics and researchers have published on a certain subject. It comprises the state of the art, encompassing significant discoveries as well as theoretical and methodological advancements on a given subject. Reviews of the literature do not present newly conducted experiments; instead, they rely on secondary sources. A literature review enables us to improve and showcase our abilities in two primary domains: locating knowledge and evaluating it critically.

2.1 INTEGRATED NATURAL SCENE TEXT LOCALIZATION AND RECOGNITION

Nowadays reading words from an unconstrained and noisy image is not easy. Text localization and recognition in an image is a research area which takes efforts to develop a computer system with an ability to automatically read the text from images. The Optical Character Recognition (OCR) tool gives good results obtained to read the text from an image. The objective of this study is to propose a new method for text localization and recognition in natural scene images with complex backgrounds. In this paper, a hybrid methodology is suggested which extracts text from natural scene images with chaotic backgrounds.

The proposed approach involves four stages. First, superimposed text regions in an image are extracted based on character descriptor features like Area, Bounding box, Perimeter, Euler number, Horizontal crossings. In the second step, superimposed text regions are tested for text content or nontext using character descriptors and SVM classifiers. In the third step, detection of multiple lines in localized text regions is done and line segmentation is performed using horizontal profiles. In the final step, using vertical profiles each character of the segmented line is extracted. The workout has been done using images drawn from ICDAR 2013 and SVT 2010 datasets

2.2 A REVIEW OF OPTICAL TEXT RECOGNITION FROM DISTORTED SCENE IMAGE

The algorithms used for scene text recognition that focus on the accuracy and consistency of scene text recognition on various common datasets and compare them. In addition, to find the weakness and inconsistencies of various scene text recognition algorithms between different datasets. A PRISMA method flow diagram applies to conduct the review. The results show Convolutional Neural Network (CNN) is the most adopted approach to creating scene text recognition programs.

The highest accuracy is the CA-FCN algorithm used for the SVT dataset. However, the consistency of algorithm performance varies from one dataset to another. Most algorithms struggled with the IC15 irregular or SVT regular dataset and performed best using the IC03 dataset.Text recognition is an important direction in the field of computer vision. With the continuous development of deep learning fields such as computer vision, pattern recognition, and machine learning, scene text recognition of deep learning has been developed on this basis. The main contributions of this paper are as follows. Firstly, for the low accuracy of recognition results, data labeled as hard labels will introduce noise and loss of information, leading to poor generalization of the model and recognition results being easily affected. Secondly, adding label smoothing to obtain soft labels, which carry more information, are more robust to noise, and improve the generalization ability of the model. Thirdly, after combining CTC with label smoothing, the loss function after label smoothing is redefined.

2.3 DEEP LEARNING FRAMEWORK FOR RECOGNIZING VERTICAL TEXTS IN NATURAL SCENE

Text localization and recognition in an image is a research area which takes efforts to develop a computer system with an ability to automatically read the text from images. The Optical Character Recognition (OCR) tool gives good results obtained to read the text from an image. The objective of this study is to propose a new method for text localization and recognition in natural scene images with complex backgrounds. In this paper, a hybrid methodology is suggested which extracts text from natural scene images with chaotic backgrounds. The proposed approach involves four stages. First, superimposed text regions in an image are extracted based on character descriptor features like Area, Bounding box, Perimeter, Euler number, Horizontal crossings. In the second step, superimposed text regions are tested for text content or nontext using character descriptors and SVM classifiers. In the third step, detection of multiple lines in localized text regions is done and line segmentation is performed using horizontal profiles.
In the final step, using vertical profiles each character of the segmented line is extracted. The workout has been done using images drawn from ICDAR 2013 and SVT 2010 datasets. The results demonstrate the effectiveness of the proposed method, which can be used as an efficient method for text localization and recognition in natural scene images.

In addition, in text recognition tasks, the DenseNet model is used to construct character recognition based on Kares. Finally, the output of Softmax is used to classify each character. Our method can replace the artificially defined features with automatic learning and context-based features. It improves the efficiency and accuracy of recognition, and realizes text detection and recognition of natural scene images. Scene-based text detection is a key area due to its adverse applications. This research focuses on scene-based text detection with the aid of YOLO based object detector and a CNN-based classification approach

2.4 A HYBRID DEEP NEURAL NETWORK FOR URDU TEXT RECOGNITION IN NATURAL IMAGES

This work presents a benchmark and a hybrid deep neural network for Urdu Text Recognition in natural scene images. Recognizing text in natural scene images is a challenging task, which has attracted the attention of computer vision and pattern recognition communities. In recent years, scene text recognition has widely been studied where; state-of-the-art results are achieved by using deep neural network models. However, most of the research works are performed for English text and a less concentration is given to other languages. This paper investigates the problem of Urdu text recognition in natural scene images. Urdu is a type of cursive text written from right to left direction where two or more characters are joined to form a word. In recent years, scene text recognition has widely been studied where; state-of-the-art results are achieved by using deep neural network models. However, most of the research works are performed for English text and a less concentration is given to other languages. In this paper, we investigate the problem of Urdu text recognition in natural scene images. Urdu is a type of cursive text written from right to left direction where two or more characters are joined to form a word. Recognizing cursive text in natural images is considered an open problem due to variations in its representation. A hybrid deep neural network architecture with skip connections, which combines convolutional and recurrent neural networks, is proposed to recognize the Urdu scene text. We introduce a new dataset of 11500 manually cropped Urdu word images from natural scenes and show the baseline results. The network is trained on the whole word image avoiding the traditional character based classification. Data augmentation technique with contrast stretching and histogram equalizer is used to further enhance the size of the dataset. The experimental results on original and augmented word images show state-of-the-art performance of the network. Similar to the recognition of captions, pictures, or overlapped text that typically appears horizontally, multi-oriented text recognition in video frames is challenging since it has high contrast related to its background. Multi-oriented form of text normally denotes scene text which makes text recognition further stimulating and remarkable owing to the disparaging features of scene text.

2.5 CURSIVE TEXT RECOGNITION IN NATURAL IMAGES

The accuracy of current natural scene text recognition algorithms is limited by the poor performance of character recognition methods for these images. The complex backgrounds, variations in the writing, text size, orientations, low resolution and multi-language text make recognition of text in natural images a complex and challenging task. Conventional machine learning and deep learning-based methods have been developed that have achieved satisfactory results, but character recognition for cursive text such as Arabic and Urdu scripts in natural images is still an open research problem.Optical character recognition (OCR) techniques proposed for Arabic and Urdu scanned documents perform very poorly when applied to character recognition in natural images. In this paper, we propose a multiscale feature aggregation (MSFA) and a multi-level feature fusion (MLFF) network architecture to recognize isolated Urdu characters in natural images. The network first aggregates multi-scale features of the convolutional layers by up-sampling and addition operations and then combines them with the high-level features. Text recognition, also known as optical character recognition (OCR), converts printed or handwritten text into a digital format that is easy to edit, search and analyze.

2.6 DEVANAGARI AND BANGLA TEXT EXTRACTION FROM NATURAL SCENE

With the increasing popularity of digital cameras attached with various handheld devices, many new computational challenges have gained significance. One such problem is extraction of texts from natural scene images captured by such devices. The extracted text can be sent to OCR or to a text- to-speech engine for recognition. In this article, it proposes a novel and effective scheme based on analysis of connected components for extraction of Devanagari and Bangla texts from camera captured scene images. A common unique feature of these two scripts is the
presence of headline and the proposed scheme uses mathematical morphology operations for their extraction. Additionally, we consider a few criteria for robust filtering of text components from such scene images. Moreover, we studied the problem of binarization of such scene images and observed that there are situations when repeated binarization by a well-known global thresholding approach is effective.We tested our algorithm on a repository of 100 scene images containing texts of Devanagari and / or Bangla. This extraction involves optical character recognition (OCR) techniques, which analyze the image to recognize and convert the text into machine-readable characters. It enables the conversion of textual information from images into editable and searchable formats, facilitating various applications such as document digitization, language translation, and content indexing The input image is preprocessed to enhance its quality and improve the OCR accuracy. This may include operations like resizing, normalization, and noise reduction.Using techniques like edge detection, contour analysis, or deep learning-based models, the system identifies regions in the image likely to contain text. This step is crucial for isolating text from other elements in the scene.The final step involves compiling the recognized characters into coherent words and sentences. The extracted text can then be used for various applications, such as language translation, text-to-speech, or indexing for information retrieval

2.7 NATURAL SCENE CHARACTER RECOGNITION USING CNN

The challenging problem of cursive text recognition in natural scene images. In particular, we have focused on isolated Urdu character recognition in natural scenes that could not be handled by traditional Optical Character Recognition (OCR) techniques developed for Arabic and Urdu scanned documents. It also presents a dataset of Urdu characters segmented from images of signboards, street scenes, shop scenes and advertisement banners containing Urdu text. A variety of deep learning techniques have been proposed by researchers for natural scene text detection and recognition. Deep learning techniques have been proposed by researchers for natural scene text detection and recognition. In this work, a Convolutional Neural Network (CNN) is applied as a classifier, as CNN approaches have been reported to provide high accuracy for natural scene text detection and recognition. A dataset of manually segmented characters was developed and deep learning based data augmentation techniques were applied to further increase the size of the dataset. The

training is formulated using filter sizes of 3x3, 5x5 and mixed 3x3 and 5x5 with a stride value of 1 and 2. The CNN model is trained with various learning rates and state-of-the-art results are achieved. The final step involves compiling the recognized characters into coherent words and sentences. The extracted text can then be used for various applications, such as language translation, text-to-speech, or indexing for information retrieval. These steps collectively enable the extraction of text written in Devanagari and Bangla scripts from natural scene images, making it accessible for further processing and analysis. Advanced deep learning models, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), play a significant role in achieving high accuracy in these tasks.

2.8 BENCHMARKING SCENE TEXT RECOGNITION IN DEVANAGARI, TELUGU, AND MALAYALAM

Inspired by the success of Deep Learning based approaches to English scene text recognition, it pose and bench-mark scene text recognition for three Indic scripts - Devanagari, Telugu and Malayalam. Synthetic word images rendered from Unicode fonts are used for training the recognition system.We use a segmentation free, hybrid but end-to-end trainable CNN-RNN deep neural network for transcribing the word images to the corresponding texts. Urdu Natural Scene Character Recognition using Convolutional Neural Networks (CNNs) involves leveraging deep learning techniques to accurately identify and recognize characters from images captured in natural scenes.Apply preprocessing techniques to enhance the quality of the images. This may involve resizing, normalization, and contrast adjustment to ensure consistency and improve the CNN's ability to learn relevant features. Design a CNN architecture tailored for Urdu character recognition. CNNs are well suited for image-related tasks due to their ability to automatically learn hierarchical features.Evaluate the performance of the trained CNN on the testing dataset to assess its ability to recognize Urducharacters in natural scenes.

2.9 AUTOMATED LOCALIZATION AND RECOGNITION OF PERSPECTIVELY DISTORTED TEXT IN NATURAL SCENE IMAGES

The automated understanding of textual information from the image is the main goal of Scene Text Recognition (STR). STR is very difficult due to several reasons such as viewing angle and lighting, which are not carefully controlled and very little amount of linguistic context contained in scene images due to which other objects present in the image can interfere with the recognition process. Most of the existing work is focused on the recognition of texts which are frontal parallel and horizontal to the image plane. We formulate a novel method to recognize text in natural scene images which are prospectively distorted. Our method uses the Hough Transform to correct the orientation of scene images and uses efficient character detection and localization techniques. Apply activation functions (e.g., ReLU) to introduce non-linearity, and use pooling layers (e.g., max pooling) to downsample and reduce the spatial dimensions of the features.Flatten the output from the convolutional layers into a vector, preparing it for input into fully connected layers. This extraction involves optical character recognition (OCR) techniques, which analyze the image to recognize and convert the text into machine-readable characters. It enables the conversion of textual information from images into editable and searchable formats, facilitating various applications such as document digitization, language translation, and content indexing

2.10 TEXT RECOGNITION SYSTEM FOR VISUALLY IMPAIRED USING PORTABLE CAMERA

Text detection in natural and complex images play an important role in analysis of images. Text information appears everywhere like product labels, documents, scene images etc. It will be very difficult for visually impaired people to find text regions in an image. This problem should be addressed because it cannot assume that the captured image contains only text. This paper presents a camera-based text reading framework

which helps visually impaired people to read texts in natural scenes, product labels etc. Here, the task is divided into two steps i.e. text detection and text recognition. Text detection involves text localization phase. Firstly, we binarize the image. Geometric and stroke width filtering removes non text regions of an image. Segmented text regions in a cluttered scene are binarized and recognized by optical character recognition. Recognized texts are output to the visually impaired people in speech. Experimental results show that the proposed method offers better performance in text recognition. The system utilizes a portable camera, commonly integrated into a smartphone or a dedicated handheld device. Users point the camera towards the text they want to read. The captured image from the camera serves as input to the system. The image may contain printed or handwritten text.Apply preprocessing techniques to enhance the quality of the captured image. This may involve resizing, noise reduction, and contrast adjustment to ensure better performance during text extraction. Use text detection algorithms to identify regions in the image that likely contain text. This step helps in isolating the text from the rest of the scene.Refine the detected text regions to precisely identify the boundaries of individual words or lines. Accurate localization is crucial for ensuring that only the relevant text is processed.Implement additional accessibility features, such as voice prompts, gestures, or vibration feedback, to enhance the user experience and provide information about the orientation of the camera or the presence of obstacles

III. EXISTING SYSTEM

The technological advancements of today's fast-paced digital world have brought about a tremendous revolution in the recruitment scene. This evolution is demonstrated by the current interview chatbot system, which marks a significant transition from traditional interview approaches to an advanced automated approach. This system is a combination of advanced Natural Language Processing (NLP) algorithms and interactive conversational interfaces, specifically designed to negotiate the difficulties of applicant assessment. It was created in response to the increasing need to optimize and streamline the hiring process to meet the demands of effective candidate screening as well as the difficulties presented by a large number of applications. The current interview chatbot system is thoroughly analyzed and evaluated in this paper, which explores its strengths, weaknesses, functions, and possible areas for improvement.

- **Google lens:** Google Lens is an image recognition technology developed by Google. It uses artificial intelligence to analyze and understand visual information taken from a mobile device's camera or images stored in the device. Google Lens can identify objects, landmarks, plants, animals, and products by scanning them with the camera, and provide relevant information and actions based on the recognized item.

- **Tesseract OCR**:An open-source OCR engine developed by Google. It's widely used for text extraction from images or scanned documents. Tesseract supports over 100 languages and it can recognize various fonts and writing styles.Tesseract OCR works by analyzing the shapes and patterns of characters in an image and interpreting them as text. It applies various

image processing techniques, such as binarization and noise reduction, to enhance the accuracy of the recognition

- **OpenCV**: A popular computer vision library that includes tools for text detection and recognition. It allows users to perform a wide range of tasks related to computer vision, such as image/video processing, object detection and tracking, feature extraction, image stitching, augmented reality, and many others. OpenCV provides a comprehensive set of functions for these tasks, including various image filters, edge detection, image segmentation, feature extraction, and machine learning algorithms

IV. LIMITATIONS IN EXISTING SYSTEM

The current interview procedures face several difficulties, such as laborious manual screenings, slow response times, a lack of real-time insights, and disjointed coordination. These challenges highlight the need for a novel interview chatbot system to overcome these challenges and improve the effectiveness of candidate assessment for improved hiring procedures.

- **Complex and diverse font styles:** Malayalam language has a wide range of font styles, each having its own unique characteristics. Existing systems may struggle to accurately recognize and interpret all these different font styles, leading to lower accuracy in text detection and recognition.

- **Image quality and noise:** Natural scene images often have variations in lighting, perspective, and image quality. This can introduce noise and blur in the image, making it difficult for existing systems to accurately identify and extract the Malayalam text

- **Occlusion and complex backgrounds:** Text in natural scene images can be occluded by other objects or have complex backgrounds. These occlusions and complex backgrounds can make it difficult for existing systems to accurately locate and extract the text from the image process.

- **Dependency on Training Data Quality**: The caliber and variety of a chatbot's training data have a significant impact on its efficacy. The chatbot's replies and decision-making may be less than ideal if the original dataset it was trained on is biased, lacks diversity, or is not properly representative of the applicant pool. This reliance on high-quality training data may limit the chatbot's flexibility in responding to different candidate profiles and needs, which would reduce its overall accuracy and usefulness in the hiring process.

- **Lack of language-specific OCR tools:** While optical character recognition (OCR) has been developed for several languages, there may be a scarcity of language-specific OCR tools for Malayalam. This can limit the availability of specialized algorithms and techniques that can be tailored specifically for Malayalam text detection and recognition process.

V. PROBLEM STATEMENT

The problem statement of scene image to text recognition is to develop a system that can accurately extract and convert the text content present in an image of a scene into machine-readable text. This includes the challenges of accurately recognizing and extracting text from images with complex backgrounds, distorted text, varying lighting conditions, different fonts, and languages. The goal is to provide a reliable and efficient solution for extracting text from scene images, which can be used in various applications such as automatic captioning, information retrieval, and enhancing accessibility for visually impaired individuals.Malayalam is the native language of the Indian state of Kerala and is the principal language spoken in the region. Although there have been advancements in text detection and recognition for other languages, such as English, there is a lack of research and resources for Malayalam. The system should be able to process images taken in various lighting conditions, angles, and distances, and accurately identify the Malayalam text present in the scene. This includes not only printed text but also handwritten text. The system should be robust enough to handle variations in text size, fonts, and styles commonly found in natural scene images. The developed system should consist of two main components: text detection and text recognition. Text detection involves localizing the regions of interest within the image that contain Malayalam text, while text recognition involves converting the detected text into digital characters that can be understood and processed by a computer. The system should have high accuracy in both text detection and recognition, as well as being fast and efficient. It should be designed with real-world applications in mind, such as automated translation systems, document analysis, information retrieval, and assistive technologies for visually impaired individuals. To overcome the lack of resources and research in Malayalam text detection and recognition, it may be
necessary to build a large-scale dataset of annotated Malayalam scene images. This dataset can be used for training and evaluating machine learning models and algorithms. The system should be accurate, robust, and efficient, and can be applied to various real-world applications.

VI. PROPOSED SYSTEM

Designing a system for scene image to text recognition involves combining various techniques for effective text detection and recognition. Here's a conceptual outline for a proposed system:

1. **Text Detection**: Utilize a robust text detection algorithm, such as EAST, TextBoxes, or FOTS, to identify text regions within the scene image.Implement a Convolutional Neural Network (CNN) for text detection.Train the model on annotated data to recognize Malayalam text regions within the bounding boxes.Employ post-processing techniques like non-maximum suppression to refine results.

2. Text Recognition: Use Optical Character Recognition (OCR) for Malayalam text recognition.Train an OCR model on a dataset of annotated Malayalam text images.Employ an OCR engine like Tesseract or a deep learning-based approach like CRNN, ASTER, or MORAN for accurate text recognition within the detected regions.Ensure the recognition model is capable of handling different fonts, languages, and text sizes commonly found in natural scenes.

3. **Image Acquisitio**n: Collect natural scene images containing Malayalam text.Ensure a diverse set of images with variations in lighting, orientation, and background.

4. Post-Processing and Refinement:Implement post-processing techniques to refine the detected text regions and improve recognition accuracy .Explore methods such as geometric verification or context-based filtering to enhance the precision of text localization.

5. **Text Localization**:Use techniques like edge detection and morphological operations to locate text regions.Extract bounding boxes around potential text regions.

6. **Language Modeling:**Employ language modeling techniques to enhance recognition accuracy.Use language models that are specifically trained on Malayalam text for better understanding and context.

7. **Offline Application**:Develop a user-friendly offline application with a graphical user interface (GUI).Allow users to upload images or capture images using a camera.Integrate the text detection and Scene Image to Text Recognition in Malayalam App Nehru College of Engineering and Research Centre 21 recognition modules into the application.

8. Deployment:Package the application for offline use on various platforms (Windows, macOS, Linux).Ensure that the application runs efficiently and provides real-time or near-real-time performance

. 9. Few-Shot Learning: Investigate few-shot learning techniques to enable the system to recognize text in new scenes with minimal labeled data, improving adaptability to diverse scenarios.

10. Privacy Considerations: Implement privacy-aware design principles, especially if the system is handling sensitive information. Consider techniques like federated learning or on-device processing to minimize data exposure

11. Dynamic Text Size Handling: Ensure the system can handle varying text sizes within scenes, adapting to scenarios where text may appear very small or large .

12. Error Analysis and Feedback Loop: Establish an error analysis mechanism to identify common mistakes and areas of improvement. Implement a feedback loop for continuous refinement based on user feedback and system performance.

13. Edge Computing for On-Device Processing: Consider deploying the system on edge devices for on-device processing, reducing latency and addressing privacy concerns by minimizing data transmission.

VII. RESULTS AND DISCUSSION

In conclusion, the development of scene image to text recognition technology in Malayalam marks a significant advancement in language processing and accessibility. This innovative application offers numerous benefits, including the ability to extract
text from images in real-time without requiring an internet connection.The integration of this technology into an offline application specifically designed for Malayalam script enables users to conveniently and accurately convert scene images into editable text.This has wide-ranging implications, particularly in improving accessibility for individuals with visual impairments, facilitating efficient data entry, and preserving historical or cultural texts in digital formats.Moreover, the ability to perform text recognition offline ensures user privacy and data security by eliminating the need to upload sensitive information to online servers. This addresses concerns related to data breaches and offers a more reliable and secure solution for users.Overall, the development of offline scene image to text recognition in Malayalam represents a commendable stride towards technological inclusivity and convenience, promising a future where language barriers are overcome and information accessibility is more seamless and universal

REFERENCES

[1] K. S. Satwashil and V. R. Pawar, "Integrated natural scene text localization and recognition," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2017, pp. 371-374, doi: 10.1109/ICECA.2017.8203708.

[2] O. O. Sumady, B. J. Antoni, R. Nasuta, Nurhasanah and E. Irwansyah, "A Review of Optical Text Recognition from Distorted Scene Image," 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), Prapat, Indonesia, 2022, pp. 1-5, doi: 10.1109/ICORIS56080.2022.10031325.

[3] Y. L. Ong, B. T. Lau, A. Chai and C. McCarthy, "A Deep Learning Framework for Recognizing Vertical Texts in Natural Scene," 2019 International Conference on Computer and Drone Applications (IConDA), Kuching, Malaysia, 2019, pp. 48-53, doi: 10.1109/IConDA47345.2019.9034915.

[4] A. Ali and M. Pickering, "A Hybrid Deep Neural Network for Urdu Text Recognition in Natural Images," 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), Xiamen, China, 2019, pp. 321-325, doi: 10.1109/ICIVC47709.2019.8980992.

[5] A. A. Chandio, M. Asikuzzaman, M. R. Pickering and M. Leghari, "Cursive Text Recognition in Natural Scene Images Using Deep Convolutional Recurrent Neural Network," in IEEE Access, vol. 10, pp. 10062-10078, 2022, doi: 10.1109/ACCESS.2022.3144844.

[6] U. Bhattacharya, S. K. Parui and S. Mondal, "Devanagari and Bangla Text Extraction from Natural Scene Images," 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 171-175, doi: 10.1109/ICDAR.2009.178.

[7] A. A. Chandio, M. Asikuzzaman and M. R. Pickering, "Cursive Character Recognition in Natural Scene Images Using a Multilevel Convolutional Neural Network Fusion," in IEEE Access, vol. 8, pp. 109054-109070, 2020, doi: 10.1109/ACCESS.2020.3001605.

[8] M. Mathew, M. Jain and C. V. Jawahar, "Benchmarking Scene Text Recognition in Devanagari, Telugu and Malayalam," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 42-46, doi: 10.1109/ICDAR.2017.364.

[9] A. Cherian and S. Sebastian, "Automatic localization and recognition of perspectively distorted text in natural scene images," 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), Pudukkottai, India, 2016, pp. 1-6, doi: 10.1109/ICETETS.2016.7602995

.[10]G. Vaidya, K. Vaidya and K. Bhosale, "Text Recognition System for Visually Impaired using Portable Camera," 2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW), Mumbai,

India, 2020, pp. 1-4, doi: 10.1109/ICCDW45521.2020.93187.

[11] H. Liu and A. Zhu, "Synthesizing Scene Text Images for Recognition with Style Transfer," 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), Sydney, NSW, Australia, 2019, pp. 8-13, doi: 10.1109/ICDARW.2019.40073