Grenze International Journal of Engineering and Technology, June Issue

Scene Text Recognition of Indian Languages in Natural Scene Images

Sanugommula Sai Pranitha[1], Durgaraju Sri Rama Charan[2], Borkut Sharan Paul[3] and R. Suvarna Rao[4]
[1-3]Student, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India
Email: pranithasanugommula@gmail.com, sriramacharan31@gmail.com, sharanpaulb24@gmail.com
[4]Assistant Professor, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India
Email: r.suvarnarao@iare.ac.in

***Abstract*— Scene Text Recognition (STR) for Telugu and Sanskrit is challenging due to complex character shapes, modifiers, and noisy backgrounds in natural images. Traditional OCR systems struggle to handle such variations effectively. This paper proposes a hybrid deep learning model that combines Convolutional Neural Networks (CNN) for spatial feature extraction with Bidirectional Long Short-Term Memory (Bi-LSTM) networks for sequence learning. The system is trained on a large labeled dataset with preprocessing techniques such as resizing and normalization. Experimental results show improved accuracy and robustness, making the approach suitable for digitization and assistive applications.**

***Index Terms*— Scene Text Recognition, Optical Character Recognition, Convolutional Neural Networks, Bi-LSTM, Deep Learning, Telugu, Sanskrit.**

I. INTRODUCTION

In recent years, Text recognition in natural scene photos has garnered a lot of interest lately because of its useful applications in a number of fields, such as automated signboard reading, digitization initiatives, and support for people with visual impairments. The process of recognizing and deciphering text embedded in intricate and cluttered backgrounds, like street signs, ads, and public displays, is known as scene text recognition. This task becomes more complex.

when working with Indian scripts, especially Telugu and Sanskrit, which are distinguished by their complex structures, many ligatures, curves, and modifiers. The development of strong recognition systems that can precisely recognize and interpret these scripts in real-world settings is essential to overcoming these obstacles. Objectives of the paper are: • Develop a scene text recognition system for Indian languages using deep learning. • Implement an end-to-end pipeline combining text detection and text recognition. • Enhance recognition accuracy for natural scene images containing complex backgrounds and multilingual content. The limitations of conventional tools like Tesseract have been brought to light by the increasing demand for optical character recognition (OCR) systems in uncontrolled environments. Even though these tools work well with printed English text, they frequently have trouble understanding the intricacies of Indian scripts, particularly when the text is handwritten, distorted, or embedded in natural images. The majority of OCR models currently in use are mainly made for English or Latin-based scripts, which means that Indian scripts like Telugu and Sanskrit are not well supported or addressed. By creating a strong scene text recognition system tailored for Telugu and Sanskrit using a hybrid deep learning architecture, this project seeks to close this gap.

Grenze ID: 01.GIJET.12.2.18

At first, we assessed conventional OCR tools such as Tesseract, but their accuracy was inadequate because of background noise, a variety of fonts, and distinctive script features. Because of these difficulties, a deep learning strategy that can extract intricate temporal and spatial features straight from the data had to be used. Conventional OCR tools work fairly well on clean, printed documents, but they can't handle variations in natural scenes or non-Latin scripts like Telugu and Sanskrit. These tools are unsuitable for real-world applications because they frequently malfunction when faced with handwritten text, background noise, and changing lighting conditions. We suggest a hybrid deep learning approach that combines Bidirectional Long Short-Term Memory (BiLSTM) networks for sequence modeling and Convolutional Neural Networks (CNNs) for feature extraction in order to overcome these drawbacks. Bi-LSTM networks efficiently model temporal dependencies in sequential or connected text patterns, whereas CNNs are excellent at capturing character-level spatial features like strokes and curves. By combining these deep learning components, our system is able to accurately recognize Telugu and Sanskrit scripts even in cluttered and noisy environments by capturing their spatial and sequential features.

Individual character images are subjected to visual feature extraction by the CNN module, and their structural patterns are sequentially interpreted by the Bi-LSTM component. This combination greatly improves recognition accuracy in a variety of scenarios, such as handwritten samples, faded prints, and background-interfering images. Additional difficulties with natural scene images include different lighting, occlusion, distortion, and overlapping objects. In these settings, text may blend with non-text elements like logos or patterns, appear at different angles, or be partially obscured. These differences make recognition more difficult for scripts with complex structures, like Telugu and Sanskrit. As a result, the scene text recognition system needs to be resilient, flexible, and able to handle variations in image layout and quality. We created a multi-stage pipeline to address these issues. The dataset first goes through preprocessing procedures like noise reduction, image resizing, normalization, and binarization. To standardize input dimensions for training, every image is uniformly resized to 32 x 32 pixels. CNN layers for feature extraction, followed by Bi-LSTM layers to interpret the sequential nature of the characters.

Due to irregular formatting and erratic backgrounds, scene text recognition in natural images—like those taken from store signs, street signs, and handwritten notes—presents considerable difficulties. Conventional OCR tools struggle with such complexities and are best suited for clean, printed content. By introducing a CNN-BiLSTMbased recognition system tailored for Telugu and Sanskrit, our project fills this gap and guarantees improved performance and dependability in scenebased character recognition tasks.

II. Literature Survey

Tesseract and other traditional optical character recognition (OCR) systems have shown promise in identifying printed Latin-based scripts, but they struggle greatly with intricate Indian scripts like Telugu and Sanskrit. These scripts have curved strokes, diacritical marks, and ligatures that make segmentation and classification more difficult. Smith (2007) points out that while Tesseract's adaptive classifier and line-finding methods were novel when they were first introduced, their performance significantly deteriorates when dealing with handwritten text or noisy images, especially when used with Indian languages. Researchers have started looking into deep learning-based methods to improve the accuracy of character recognition in order to get around these restrictions. [1]

Convolutional Neural Networks (CNNs), in particular, are adept at automatically extracting hierarchical features from unprocessed input data, as noted by LeCun et al. (2015). Because deep CNNs can learn spatial patterns, they are especially good at image processing tasks like character recognition in noisy and complex scene images. When used for OCR in Indian scripts, CNNs have produced better results than conventional techniques. However, sequence modeling remains difficult for CNNs alone, particularly when it comes to identifying character sequences in scene text. [2].

CNNs and Recurrent Neural Networks (RNNs), like Long Short-Term Memory networks (LSTMs), have been combined in recent years. By facilitating the extraction of both spatial and temporal features, this hybrid CNN-RNN model greatly increases recognition accuracy. By developing highperformance architectures for image and sequence respectively, Krizhevsky et al. (2012) and Hochreiter & Schmidhuber (1997) laid the groundwork for these models. These models improve overall recognition rates across different scripts because they are better able to capture the character flow in scene text images. [3]

In their thorough investigation of OCR systems for Indian scripts, JJain et al. (2021) found several significant issues, such as low-resolution scans, a variety of handwriting styles, and a lack of datasets. They noted that many Indian script OCR models still depend on manually crafted features or shallow learning techniques. In order to

improve accuracy and flexibility, the paper promotes the use of deep learning architectures that have been trained on sizable, varied datasets. To improve recognition quality under different image conditions, our system uses CNN-based feature extraction in conjunction with preprocessing methods like binarization, noise reduction, and morphological operations. [4]

Deep learning models were developed especially for the Bangla, Hindi, and Devanagari scripts—which are structurally similar to Telugu and Sanskrit—by Saha et al. (2020) and Kumar & Subramanian (2019). By using data augmentation techniques to increase model robustness and training on multilingual datasets, these models achieved state-of-the-art performance. For efficient OCR performance, their research emphasized the importance of script-specific tuning, careful preprocessing, and the integration of various data sources. Our goal is to use Convolutional Neural Networks (CNNs) to create a reliable and scalable system for Telugu and Sanskrit character recognition. [5]

III. PROPOSED METHOD

The objective of this project is to use deep learning to identify Telugu and Sanskrit characters from photos of natural scenes. In order to identify complex scripts like Telugu and Sanskrit, we first experimented with standalone Convolutional Neural Networks (CNNs), which are good at extracting spatial features like edges, curves, and patterns. However, sequential data—such as linked handwritten characters or flowing text in images— presents difficulties for CNNs alone. We used a hybrid model that combined CNN and Bi-directional Long Short-Term Memory (Bi-LSTM) networks to get around this. In this configuration, Bi-LSTMs record the temporal and contextual relationships between characters, while CNNs extract spatial features.

For training, we used a custom dataset of handwritten and printed Telugu and Sanskrit characters collected from scanned documents, digital prints, and user-submitted samples. This dataset contains over 290,000 images representing 631 unique characters, spanning multiple fonts, styles, and script combinations. All images were normalized to a uniform 32×32 resolution to maintain consistency across inputs. The dataset also includes annotated scene text images for Indian scripts, making it suitable for training and evaluating character recognition models like ours.

In below screen showing CNN and LSTM model details used to detect scene text

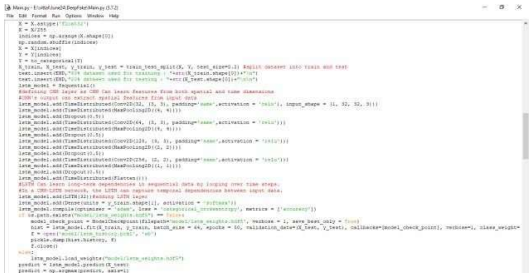

Fig.1 Combining CNN and LSTM layer

In above screen you can see we have combine both CNN and LSTM layers and can read red colour comments to know the usage of both layers.

To detect Scene Text we have downloaded images from below URL

https://github.com/gokulkarthik/Indian- Scene-Text-Dataset

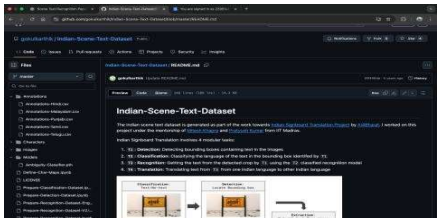

Fig.2 Scene Text

In above URL screen you can see Scene Text is recognized in the real text and this model is successfully predicting these text.

We have designed the to implement this project, following modules:

- Dataset Upload and Preprocessing: This module allows users to upload a dataset of Telugu and Sanskrit character images into the application. The application reads all the images, resizes them to a fixed 32×32 dimension, and prepares the X (features) and Y (labels) arrays for training. The dataset includes both handwritten and printed characters to ensure diversity and robustness.
- Train CNN and Bi-LSTM Model: This module normalises and shuffles the dataset, then divides it into an 80:20 train- test split. 80% of the images are used to train the CNN and Bi-LSTM models, while 20% are used for testing and validation. CNN extracts spatial features, and Bi-LSTM captures the sequential patterns in the character structure, enhancing recognition accuracy.
- Telugu and Sanskrit Character Recognition: This module allows users to upload a test image containing printed or handwritten Telugu or Sanskrit characters. The trained models analyze each segmented character and predict its corresponding label. After processing, the module provides the recognized characters as output, demonstrating the model's effectiveness in real-world scenarios.

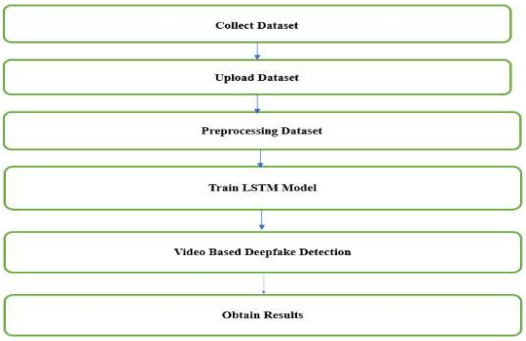

Fig.3 Flowchart of proposed method

IV. RESULT

To run the project, double-click on the run.bat file to display the screen below

Fig.4 Upload Scene Text Dataset

In above screen click on 'Upload Language Dataset button to upload dataset and get below output

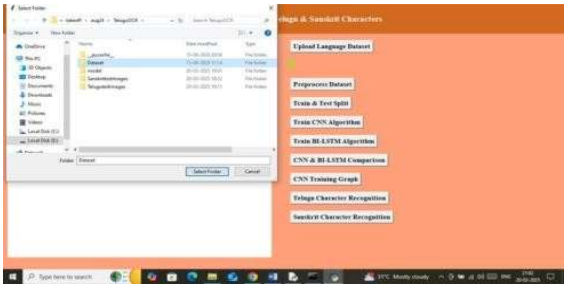

Fig.5 selecting and uploading dataset annotation file

In above screen selecting and uploading dataset folder and then click on 'Select Folder' button to load dataset and get below output.

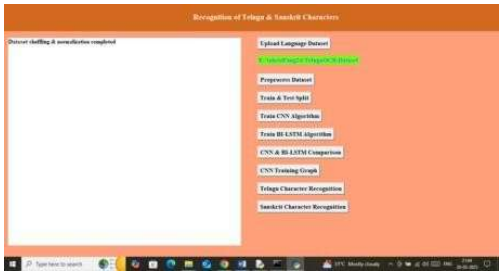

Fig. 6 train and test dataset size

In above screen dataset processing completed and now click on 'Train & Test Split' button to split dataset into train and test and get below output

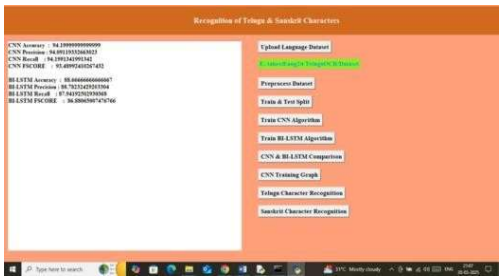

Fig.7 image analysis started

In above screen can see CNN model training completed and it got validation accuracy as 94%on testing data and can see other metrics like precision, recall and FSCORE. Now click on'Train BI-LSTM Algorithm' button to train BILSTM algorithm and then will get below output

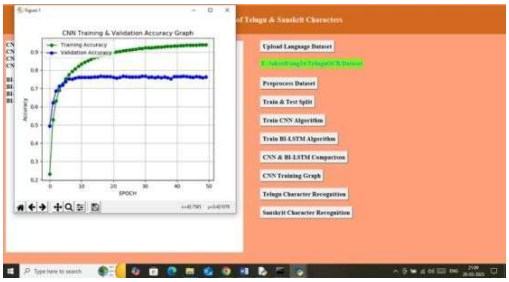

Fig. 8 gives BILSTM and CNN accuracy

In above screen BILSTM got 88% accuracy and now click on 'CNN & BI-LSTM Comparison' button to get below graph

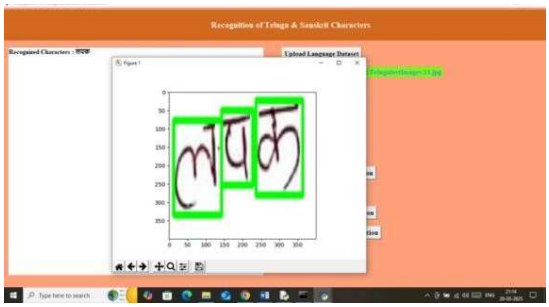

Fig. 9 delivering the output translation

In above screen can recognized characters and similarly you can upload and test other images.

V. CONCLUSION

In conclusion, the proposed Convolutional Neural Network (CNN) and Bi-directional Long Short-Term Memory (Bi-LSTM) based scene text recognition system effectively addresses the challenge of identifying complex Telugu and Sanskrit characters from natural scene images. The system automates dataset preprocessing, including normalisation, resizing, and shuffling, ensuring consistent input quality. An 80:20 train-test split is employed to robustly assess model accuracy across diverse handwritten and printed samples. By combining CNN for spatial feature extraction with Bi-LSTM for capturing sequential character patterns, the model provides precise and reliable recognition of structurally rich Indian scripts. The system also allows users to upload real-world images and receive accurate text output, even in cases of noise, distortion, or overlapping characters. This hybrid architecture significantly enhances recognition performance, especially for natural and historical document images.

REFERENCES

[1] Smith, R. (2007). An Overview of the Tesseract OCR Engine. In Proceedings of the Document Analysis and Recognition (ICDAR), IEEE.

[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.This paper introduces the fundamentals and breakthroughs of deep learning, highlighting its impact on AI and computer vision.

[3] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8),1735–1780. This foundational paper introduces the LSTM architecture, enabling neural networks to learn long-term dependencies in sequential data.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems.

[5] Nayak, P., & Kumar, S. (2019). Deep Learning for Telugu Handwritten Character Recognition. Springer, Smart Innovation, Systems, and Technologies.

[6] Sharma, M., & Gupta, R. (2018). Recognition of Sanskrit Characters Using Convolutional Neural Networks. In Proceedings of the International Conference on Computer Vision and Image Processing (CVIP)..

[7] Jain, A., Sharma, A., & Baghel, P. (2021). Optical Character Recognition (OCR) for Indian Scripts: A Survey. Journal of Emerging Technologies and Innovative Research (JETIR).

[8] Dutta, K., Krishnan, P., Mathew, M., et al. (2018). Improving Indic Script OCR Using Deep Learning. Asian Conference on Computer Vision (ACCV).

[9] Google Research Team. (2020). OCR for Indian Languages: An Open- Source Benchmark. arXiv:2012.06752..

[10] Your Project Report: Recognition of Telugu & Sanskrit Characters in Natural Scene Imagesusing CNN and Bi-LSTM. (Unpublished B.Tech Project Report, 2025).