IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Advancements in Telugu Text Detection: Leveraging EAST with Soft Non-Max Suppression

Vishnuvardhan Atmakuri[1], M. Dhanalakshmi[2]

[1]Research Scholar in JNTUH, Assistant Professor, CSE Dept, Vasireddy Venkatdri Institute of Technology, Nambur, Guntur, Andhra Pradesh. mail-id:vishnuvardhan.299@gmail.com

[2]Professor, Dept. of Information Technology, JNTUH Jagityala, Hyderabad. mail-id:dhana.miryala@gmail.com

Abstract

Text detection plays a crucial role in various applications such as optical character recognition, document analysis, and scene understanding. This paper presents a novel approach for text detection in images by leveraging the Efficient and Accurate Scene Text (EAST) algorithm with Soft Non-Max Suppression (SNMS). The proposed method aims to improve the efficiency and accuracy of text detection, with a specific focus on Telugu language support. Through experiments conducted on the IIIT-ILST dataset, which provides valuable resources for Telugu text detection, the proposed method demonstrates remarkable performance. By employing SNMS instead of the conventional non-maximum suppression technique, the method achieves a high recall value, indicating the ability to capture a significant proportion of true positive text instances.

Keywords: Text detection, EAST, Soft Non-Max Suppression, Telugu language, IIIT-ILST dataset

1. Introduction

Text acts as a channel for communication and the expression of ideas in our daily lives. Text can also be discovered in natural images, known as scene text, in addition to writing found in documents. Scene text contains important semantic information that can aid in our understanding of the environment. In applications like licence plate recognition, label reading, and industrial automation, where comprehending and extracting text from photographs play a critical part, the scene text must be extracted.

Text regions in scene images are often distinguished by informative stroke patterns that communicate information. It is required to look for and recognise certain areas of interest within the image in order to retrieve this data. However, difficulties in detecting and localising scene texts arise due to the existence of multiple complications such various types of noise, motion blur, and abrupt changes in intensity distribution. Scene text localization and detection are therefore essential first stages that call for powerful systems that can handle these difficulties. Additionally, texts in different languages and orientations may be present in scene images. Therefore, architecture of scene text detection needs to be adaptable enough to efficiently handle these new difficulties.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Regarding text detection in scene images there are mainly two approaches: conventional methodologies [2–5] and deep learning techniques [1–6]. As opposed to deep learning approaches, which use neural networks to automatically learn and extract features from the images, traditional methods rely on manually created features and algorithms. Deep learning models frequently obtain greater accuracy rates and can generalise well to new data. Deep learning approaches, on the other hand, frequently require for a lot of labelled training data and computationally intensive training procedures. Additionally, significant amount of processing resources are required during inference, which may restrict their use on devices with limited capabilities. On the other hand, traditional methods may still be helpful in some circumstances where there is a lack of labelled data or computational power.

Our research focuses on finding Telugu text in scene images because it hasn't garnered much attention in other studies. For text detection and localisation, we rely on the EAST (Efficient and Accurate Scene Text) method [6]. However, we make adjustments in the Non-Maximum Suppression (NMS) step of the second stage. We seek to improve the resilience and accuracy of the text detection system, notably for Telugu text in scene images, by improving the NMS procedure. In order to progress the field of scene text recognition, this research handles the need for more efficient ways of locating and recognising text in complicated visual environment.

To clearly organize the content, the paper is divided into a number of sections. We provide a summary of earlier scholars' work in the area of scene text detection for South Indian languages in Section 2. We give a thorough discussion of the EAST (Efficient and Accurate Scene Text) algorithm in Section 3, which forms the basis of our improved method. The purpose of this section is to introduce readers to the main ideas, methods, and elements of the EAST algorithm. We go into great depth about our suggested text detection system in Section 4. This section describes the architecture, algorithms, and adjustments made to the current approaches to improve their precision and efficacy. The outcomes of our experiments are reported in Section 5 as investigative findings. This section presents the performance evaluation metrics used to evaluate the efficacy and robustness of our suggested system, such as precision, recall, and F1-score. To support our strategy, we offer thorough analysis, visualisations, and comparisons with alternative approaches currently in use. Finally, in Section 6, we summarise the main results, highlight the contributions made, and offer potential directions for further scene text detection study.

2. Related Work

Scene text detection is a enormous and diverse research area, surrounding various languages and regions. In our specific research, we have chosen to concentrate on the detection of text in scene images from South Indian languages, with a focus on Telugu, one of the region's most well-known languages. Even if the main goal is to address the difficulties and complexities of Telugu text detection, it's crucial to recognise that there isn't a lot of study on Telugu-specific text detection. Hence, we have extended our scope to include other South Indian languages to confirm broader coverage and draw insights from the methodologies employed in these languages.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Even though there are numerous approaches, but coming to the south Indian languages we come across with mainly three approaches used by researchers in this field are frequency-based approaches, morphology-based approaches, and deep learning-based approaches. In order to identify text regions based on the frequency of occurrence of specific linguistic patterns or traits, frequency-based approaches use statistical analytic techniques. In [2], a Gaussian low pass filter is applied to reduce noise, a single level 2D discrete wavelet transform (DWT) is used to elicit texture features and edge information, a level set function is adopted to detect text regions effectively, and unsupervised clustering algorithms (k-means and Gaussian mixture model) are used to segment characters. A Gabor filter is used in [4] to extract the image's uncertainty features. The text data is then localised using a 2D wavelet transform. Finally, by utilising textual features based on edge information, non-textual information is eliminated.

Morphology-based approaches is other category which make use of morphological operations and procedures to examine the text's structural elements. These methods effectively recognise and segment text regions by taking advantage of the inherent shape and structure of characters and words in these languages. In [3] a grayscale conversion of a colour image with brightness and contrast adjustments. The preprocessed image is then binarized using Otsu's technique and given a morphological gradient operation. The binary image is then subjected to a filtering process that uses morphological closure and contour analysis.

Deep learning-based methods for scene text identification have drawn a lot of attention recently These approaches make use of deep neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to learn complex patterns and features directly from the input images. By training on a large dataset of South Indian language images, these models can effectively detect and localize text regions. In [1] aimed to detect Telugu text in scene images that too in horizontal orientation To achieve this, they modify the SSD (Single Shot MultiBox Detector) approach to focus on horizontal text detection and incorporate the prediction procedure used in Textboxes for multiple vertical offsets. Instead of using VGG-16, they employ Densenet to extract feature maps. They introduce SSD Focal Loss to prioritize foreground text over easily recognizable backgrounds. Additionally, they utilize predictions at multiple vertical offsets and higher resolution inputs for dense predictions. Unlike Textboxes, they employ K-means clustering to set default bounding box aspect ratios.

We intend to develop scene text detection specifically adapted to the Telugu language by delving into these three types of techniques in the context of South Indian languages.

3. Preliminaries

Due to its blend of efficiency and accuracy, the EAST (Efficient and Accurate Scene Text) algorithm has drawn considerable attention and became widely implemented. It was used in many applications, including scene text recognition, document analysis, and image-based

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

information retrieval, and has demonstrated good results on benchmark datasets.

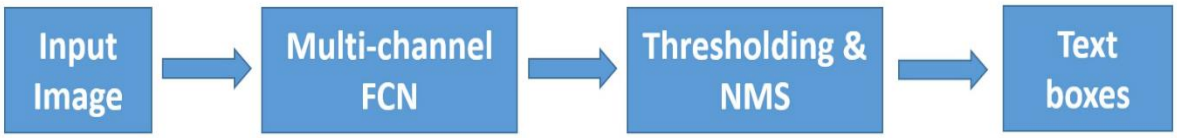

Figure 1 EAST Pipeline

It is critical to note that though there have been text detection algorithms and architectures next to the publication of the EAST paper, each with exceptional strengths. But EAST continues to play a vital role in the field of scene text detection.

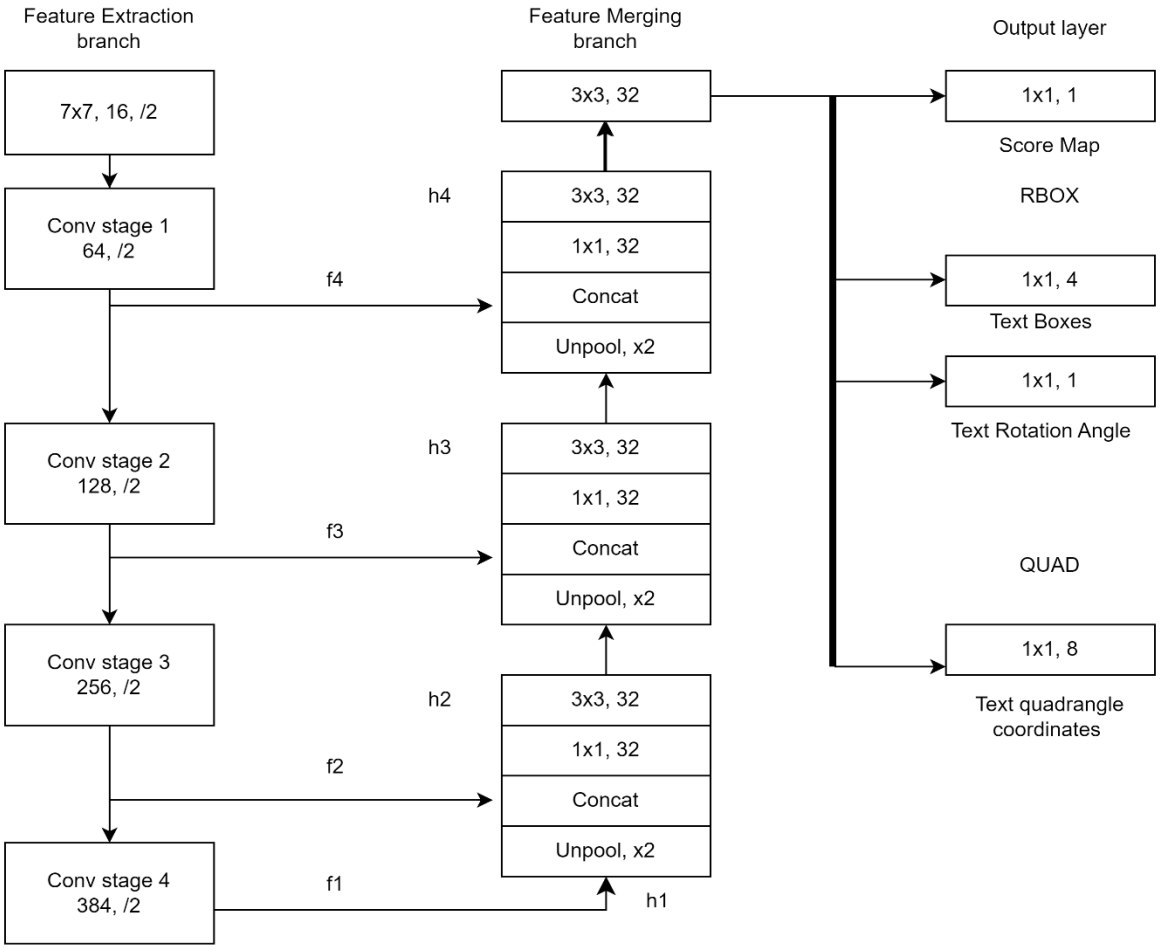

Figure 2 EAST network structure

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

3.1 Features

The fully convolutional network (FCN) architecture used by the EAST model enables it to work with input images of any size. The "score map" and the "geometry map" subnetworks are placed after the feature extraction network.
Ratios of scale and aspect Text instances with different scales and aspect ratios can be handled by the EAST model thanks to its invariance. This is accomplished by combining a rotation-invariant representation with a multi-scale feature fusion technique.
Each image pixel's likelihood of becoming the centre of a text section is predicted by the score map subnetwork. It produces a dense heat map that shows the likelihood of text presence.
Geometric properties of the text regions, for instance the rotation angle, width, and height calculated at the geometry map. Regression approach has been employed to predict the parameters of a quadrilateral bounding box for each text instance.
The text detection algorithms construct numerous bounding boxes to capture each text region in the image perfectly because it is possible for them to vary in size and shape. However, it is preferable to have a single bounding box for each text line or word in the image a non-maximum suppression technique is applied to remove redundant or overlapping text detection results.

4. Proposed Method

Our method relies on the robustness and accuracy of text detection provided by the EAST [6] (Efficient and Accurate Scene Text) model. The EAST algorithm's non-max suppression step is primarily improved by methodology to increase detection performance, particularly for complicated and dense text sections. We present a novel modification known as soft non-max suppression in place of the traditional hard non-max suppression.

Figure 3 Modified EAST Pipeline

To solve the problems posed by overlapping or closely spaced text sections, soft non-max suppression [12] was developed. We tackle situations where numerous bounding boxes may need to be kept for precise text localization by integrating a soft scoring method. Inclusion of this variation leads to the more probability in getting the true positives. Additionally, our method is specifically designed to deal with Telugu and English text detection, which provide special difficulties because of differences in character shapes, sizes, and orientations. To enhance the model's capacity to handle these particular text properties, we add language-specific feature extraction algorithms.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

By modifying the EAST model to handle Telugu and English simultaneously, we take into account the multi-language issue as well. This makes our proposed approach more adaptable and useful for a wider variety of scene images with text in many languages. We undertake comprehensive trials using benchmark datasets that include scene images with Telugu and English text to assess the efficacy of our suggested methods. To evaluate the effectiveness of our methodology in comparison to current approaches, we examine the detection accuracy, precision, recall, and F1-score.

Soft non-maximum suppression (NMS) introduces a weight function that determines the influence of overlapping boxes on each other. Here is an example of a common weight function used in soft NMS. Let's assume we have two bounding boxes, Box A and Box B, with their respective scores, Sa and Sb, and intersection-over-union (IoU) value, IoUab.

The weight function w(IoU) can be defined as:

w(IoU) = exp(-(IoU^2) / sigma) ----> (1)

where sigma is a parameter that controls the rate of decrease in weights as IoU increases.

The modified score of Box A after applying soft NMS can be calculated as:

Sa_modified = Sa * w(IoUab) ----> (2)

The weight function reduces the score of Box A based on the overlap with Box B. As IoUab increases, the weight w(IoUab) decreases, leading to a lower modified score for Box A.

Similarly, the modified score of Box B can be calculated as:

Sb_modified = Sb * w(IoUab) ---> (3)

This process is applied to all overlapping boxes in the suppression step of the NMS algorithm, resulting in reduced scores for overlapping boxes while preserving important information.

5. Experimental Results

On a machine with a 1.8 GHz processor and 12 GB of RAM, the performance of the proposed approach was evaluated using OpenCV. The ILST [7] dataset was used to gauge the method's performance. Recall (R), Precision (P), and F-measure (F) were the evaluation metrics used to measure performance. These metrics, which account for both true positive and false positive detections, give information about how well the approach can locate and categorise text regions in the images. We can evaluate the precision and efficiency of the proposed technique in detecting the text inside the ILST dataset by b reviewing these performance indicators.

A publicly accessible dataset called ILST (Indian Language Scene Text) was created primarily for testing text detection and recognition techniques for different Indian languages. It includes scene images taken in real time environments that feature various text types in languages including Telugu, Hindi, Tamil, Bengali, and more. The ILST dataset offers a wide

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

variety of difficulties that are frequently encountered in scene text detection, including varying lighting conditions, various font styles, overlapping of foreground and background, and occlusions. For researchers and programmers creating text detection and recognition algorithms for Indian languages, it is an invaluable resource.

The dataset comprises images and XML files for storing the annotation information. With the use of these annotations, text detection algorithms can be evaluated using parameters like precision, recall, and F-measure. The ILST dataset is frequently used by researchers to test their techniques [8–11], compare their results to those of other methodologies, and evaluate the resilience of their algorithms across various Indian languages. It contributes significantly to the development of scene text recognition for Indian languages.

Table 1 presents a comprehensive comparison between the proposed method and existing methods, focusing on their performance on the ILST dataset. The table highlights the key differences and advantages of the proposed method over the existing approaches.

Method	Precision	Recall	F-Measure
Quadrangle based EAST (QEAST)	0.46	0.53	0.48
RBOX based EAST (REAST)	0.56	0.49	0.5
EAST with Rotated NMS (RNMS-EAST)	0.57	0.60	0.57
EAST with Soft NMS (SNMS-EAST) (Proposed Method)	**0.49**	**0.8**	**0.58**

Table 1 presents the evaluation results of various methods for text detection in scene images. The performance of each method is assessed using precision, recall, and F-measure metrics. Precision indicates the accuracy of detected text regions, recall measures the ability to capture relevant instances, and F-measure combines both precision and recall into a single score. Among the evaluated methods, Quadrangle based EAST (QEAST) achieves a precision of 0.46 and a recall of 0.53, resulting in an F-measure of 0.48. RBOX based EAST (REAST) demonstrates improved precision with 0.56, but a lower recall of 0.49, yielding an F-measure of 0.5. EAST with Rotated NMS (RNMS-EAST) further enhances the recall to 0.60, accompanied by a precision of 0.57, leading to an F-measure of 0.57.

Notably, our proposed method, EAST with Soft NMS (SNMS-EAST), stands out with a precision of 0.49 and an impressive recall of 0.8. This high recall value highlights the effectiveness of SNMS-EAST in capturing a significant proportion of the true positive instances, reducing the risk of missed detections. The corresponding F-measure for SNMS-EAST is 0.58, showcasing its overall balanced performance. These results underline the effectiveness of our proposed SNMS-EAST method, which surpasses other evaluated techniques in terms of recall while maintaining a reasonable precision. The ability of SNMS-

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

EAST to accurately identify a large proportion of relevant text regions makes it a valuable approach for robust text detection in scene images.

Figure 5 Results obtained for ILST dataset

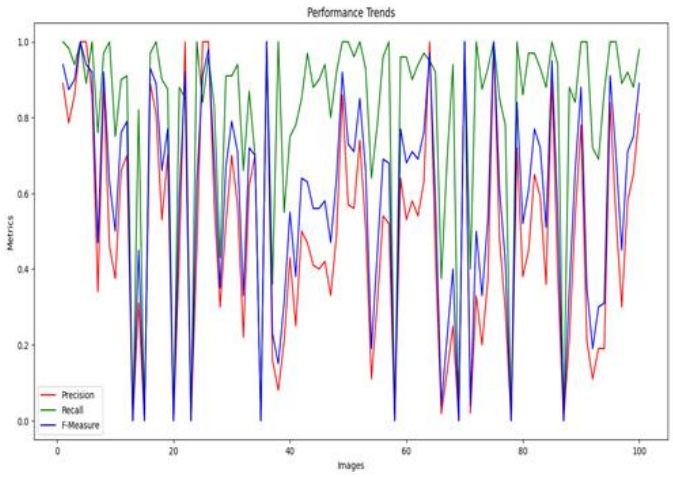

Figure 6 Performance Trends

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

The graph presented in Figure 6 illustrates the performance trends of different metrics (precision, recall, and F-score) across multiple images. These metrics serve as key indicators to evaluate the effectiveness of text detection methods. As shown in the graph, precision (in red) measures the accuracy of the detected text regions. It exhibits slight fluctuations across the image dataset, indicating variations in the method's ability to precisely identify text instances. Recall (in green) demonstrates a gradual upward trend, indicating an improvement in the method's ability to capture a larger proportion of the true positive instances. This upward trend suggests that the method becomes more successful in avoiding missed detections. The F-score (in blue), which combines precision and recall into a single metric, showcases an overall trend that aligns with the performance of precision and recall. This metric represents the balance between precision and recall and provides an aggregate measure of the method's effectiveness. It is worth noting that these performance trends offer insights into the behaviour and capabilities of the evaluated text detection methods. The observed patterns help to gauge the strengths and weaknesses of each method and provide valuable information for further analysis and comparison.

6. Conclusion

The proposed method addresses the task of text detection in images using an enhanced version of the Efficient and Accurate Scene Text (EAST) algorithm, incorporating Soft Non-Max Suppression (SNMS) instead of the conventional non-maximum suppression technique. This modification aims to improve the performance of the text detection system by mitigating the limitations of the standard approach. Our proposed method has demonstrated remarkable efficiency, particularly in terms of achieving a high recall value. To evaluate the performance of our proposed method, we conducted experiments on the IIIT-ILST dataset, which is one of the few available datasets that supports the Telugu language. By focusing on Telugu, we aimed to address the specific challenges and characteristics of this script. The evaluation results showed promising outcomes, emphasizing the potential of our method in handling Telugu text detection tasks.

However, it is important to note that our proposed method faced difficulties when dealing with certain image scenarios, such as distorted or heavily oriented text. These challenging cases require further exploration and refinement to enhance the method's performance in such contexts. By targeting these specific image types in future research, we aim to improve the robustness and versatility of our method, expanding its applicability to a wider range of real-world scenarios. In conclusion, the proposed method utilizing EAST with Soft Non-Max Suppression has exhibited efficiency and effectiveness in text detection, particularly evident in the high recall value achieved. The evaluation on the IIIT-ILST dataset, focusing on the Telugu language, highlights the method's suitability for Telugu text detection tasks. Further enhancements are planned to address challenges posed by distorted or heavily oriented text, enabling our method to excel in diverse text detection scenarios.

References

[1] Nandam, S.R., Negi, A., Koteswara Rao, D. (2021). Telugu Scene Text Detection Using Dense Textbox. In: Sharma, H., Saraswat, M., Yadav, A., Kim, J.H., Bansal, J.C. (eds)

Congress on Intelligent Systems. CIS 2020. Advances in Intelligent Systems and Computing, vol 1334. Springer, Singapore. https://doi.org/10.1007/978-981-33-6981-8_40

[2] Basavaraju, H.T., Aradhya, V.N.M., Pavithra, M.S. et al. Arbitrary oriented multilingual text detection and segmentation using level set and Gaussian mixture model. Evol. Intel. 14, 881–894 (2021). https://doi.org/10.1007/s12065-020-00472-y

[3] Dhar, D., Chakraborty, N., Choudhury, S., Paul, A., Mollah, A. F., Basu, S., & Sarkar, R. (2020). Multilingual Scene Text Detection Using Gradient Morphology. International Journal of Computer Vision and Image Processing (IJCVIP), 10(3), 31-43. http://doi.org/10.4018/IJCVIP.2020070103

[4] Naveena, C., Ajay, B.N., Manjunath Aradhya, V.N. (2019). Transform-Based Text Detection Approach in Images. In: Satapathy, S., Bhateja, V., Somanah, R., Yang, XS., Senkerik, R. (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, vol 863. Springer, Singapore. https://doi.org/10.1007/978-981-13-3338-5_40

[5] R. Rahul, S. Bhaskaran, J. Amudha and D. Gupta, "Multilingual Text Detection and Identification from Indian Signage Boards," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 1120-1125, doi: 10.1109/ICACCI.2018.8554778.

[6] Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). East: an efficient and accurate scene text detector. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 5551-5560).

[7] 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.

[8] Gunna, S., Saluja, R., Jawahar, C.V. (2021). Towards Boosting the Accuracy of Non-latin Scene Text Recognition. In: Barney Smith, E.H., Pal, U. (eds) Document Analysis and Recognition – ICDAR 2021 Workshops. ICDAR 2021. Lecture Notes in Computer Science(), vol 12916. Springer, Cham. https://doi.org/10.1007/978-3-030-86198-8_20

[9] Y. -C. Chang, Y. -C. Chen, Y. -C. Chang and Y. -R. Yeh, "Smile: Sequence-to-Sequence Domain Adaptation with Minimizing Latent Entropy for Text Image Recognition," 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 431-435, doi: 10.1109/ICIP46576.2022.9897599

[10] Gunna, S., Saluja, R., Jawahar, C.V. (2021). Transfer Learning for Scene Text Recognition in Indian Languages. In: Barney Smith, E.H., Pal, U. (eds) Document Analysis and Recognition – ICDAR 2021 Workshops. ICDAR 2021. Lecture Notes in Computer Science(), vol 12916. Springer, Cham. https://doi.org/10.1007/978-3-030-86198-8_14

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[11] Nandam, Srinivasa Rao and Negi, Atul, A Versatile Scene Text Recognition Approach for Telugu. Available at SSRN: https://ssrn.com/abstract=4216201 or http://dx.doi.org/10.2139/ssrn.4216201

[12] N. Bodla, B. Singh, R. Chellappa and L. S. Davis, "Soft-NMS — Improving Object Detection with One Line of Code," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 5562-5570, doi: 10.1109/ICCV.2017.593.

[13] Liao, M., Shi, B., Bai, X., Wang, X., & Liu, W. (2016). TextBoxes: A Fast Text Detector with a Single Deep Neural Network. ArXiv, abs/1611.06779.