International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Recognition of Printed Kannada Text in Scene Images using Machine Learning Techniques

Mahadeva Prasad Y. N.*[1], Dr. Chethan H. K.[2]

Submitted: 05/10/2023 **Revised**: 22/11/2023 **Accepted**: 03/12/2023

Abstract: Images are essential and plays a key and an important function in an electronics and communication media for sharing the information. In recent days, every activity has to be recorded as a digital image. Despite the rapid progress in the growth of strong deep learning representations for detecting and recognizing scene text, current state-of-the-art techniques fall short of delivering satisfactory outcomes in complex scenarios, such as those involving diverse logos, decorated boards, or text embedded within components. This proposed research article provides a new approach to enhancing the detection of scene text and recognition of scene text performances by detecting defects in the detection results. In this research, four important stages such as pre-processing, Kannada text detection, feature extraction and recognition are carried after the collection of multilingual scene text images. Initially, the multilingual scene text images are collected and is preprocessed for the effectual removal of noise using bilateral filter and wiener filter for enrich the image quality. After pre-processing, detection step is performed for analyzing only Kannada text by leaving the other texts using Convolutional block attention-based YOLO V7. The highly discriminative feature/s are picked out from the detected scene images by using Dense stacked LSTM network model. Finally, the Convolutional Residual network assisted auto encoder model is employed for optimal recognition of Kannada language can be created using machine-readable strings. This yields the better recognition rate compared to cutting-edge techniques.

Keywords: *Filters, Kannada text, Text Recognition, LSTM Model, YOLO V7*

1. Introduction

Text, as a physical representation of language, is widely regarded as one of humanity's most influential inventions. It is essential in human life because it allows people to communicate ideas, deliver knowledge, and so on. In reality the text or textual

cues are used to interpret and pass down ancient civilization. Text found in images can be of great semantic importance for the overall image understanding as a rich and precise information carrier. This technology shows an essential part in various applications, including autonomous navigation, document analysis, content indexing for search engines, and augmented reality systems. Extracting text from images has emerged as a popular and demanding research area within computer vision in recent times. Notable improvements have been implemented in text recognition from scene images. A variety of deep learning steps have been developed by scholars to detect text in natural scenes. In contrast, the Kannada language remains relatively underexplored.

Text stands as a crucial source of information for humans [1]. Scene text recognition (STR) pertains to the recognition of text in natural scene images. During the early era, there was a growing demand for a greater challenging task—extracting text from images gathered in real-world scenarios. Text embedded within images can emerge in various scenarios, including handwritten notes, official documents, menu displays, street views featuring elements like street names, advertisements, vehicle registration plates, in addition to the names of restaurants and coffee shops, among other contexts. Although there are situations where text is deliberately highlighted, such as in scanned documents or photographs of restaurant menus, it can also occur in unaltered images without a specific emphasis from the image's creator. Recognizing text is a intriguing task, encompassing the detection of text with various fonts, styles, and diverse background noise. The complexity is further heightened when dealing with handwritten text, given variations in letter size, orientation, and spacing between letters, which differ among individuals. Consequently, here is a demand for the creation of an automated text recognition system capable of identifying text components in vision or scene and converting them into a machine-recognizable format [2].

Text detection specifies the approach of identifying the positions of all text occurrences within a given image, while text recognition involves extracting machine-readable characters from a suitably cropped image containing text. In the early decade, the revelation of text in images of natural scenes has grown into a fascinating field of study in computer vision, resulting in expressive curiosity from both industry and academic research [3]. Figure 1 displays a selection of scene images captured in different locations.

[1] *Department of computer Science , University of Mysore, Mysore- 570005, Karnataka, India*
[2] *Department of computer Science and Engineering, MITT Mysore-571302 VTUniversity,Karnataka, India*
** Corresponding Author Email: mahadevprasadyn@gmail.com*

Fig. 1. Some sample scene text images

This paper examines distinct stages within the text detection and recognition process, conducting an analysis and comparison of diverse approaches employed at each stage. It examines the significance of each process and provides insights into the advantages, disadvantages, and applications of various methods employed by different contributors to address these challenges. Additionally, the paper reviews various applications of text detection and recognition.

This paper explores a critical aspect—namely, the identification and extraction of text from natural scenes. The investigation addresses challenges such as cluttered environments, diverse fonts and styles, varied backgrounds, unstructured scenes, different orientations, ambiguities, and more. The proposed methodology comprises five steps: data acquisition, pre-processing, scene text detection, feature extraction, and scene text recognition.

1.1 The Challenges of Text Detection

Recognition of scene text is for identifying the text area location and region transform of the text from the scene image into the readable or identifiable character or word or sentence. The recognition of kannada text from the natural scenes involves various constraints and challenges because of various distortions of perspective and huge curves, distinct styles in text, diversity in scale, confusing characters in the text, background complexity, low image resolution, uneven or bad luminance, image occlusions etc. [4]. Nonetheless, extracting text from images poses a formidable challenge. The primary obstacles in text detection can be classified into three groups: i) Diversity in natural images: Characters in natural images vary in font style, size, colour, and alignment. ii) Background complexity: Natural image backgrounds can be intricate, incorporating elements like grass, bricks, rocks, and signboards, adding complexity to text identification. iii) Inference factors: Key inference challenges include blurring, noise, and low resolution in input images.

2. Literature Survey

This survey also reviews and summarizes widely employed datasets that are frequently used in literature. Vishnuvardhan and Dhanalakshmi Miryala [5] presented to discover the approaches which are adapted to localize and realize the language contents of India in natural scene images. This article showcases the diverse languages of India along with their script representations. The comprehensive recognition of text in Indian languages from scene images can be attained through a camera or a smartphone. In this presentation, there is some necessary for working much more on the work by implementing a much stronger deep learning approach.

In reference [6], the authors introduce a technique for detecting and recognizing text in scene images. The paper delves into the challenges posed by various constraints and explores the future potential of research in scene text detection and identification. While many algorithms have primarily focused on English text recognition because of its relatively small alphabet set, languages like Japanese and Chinese present larger symbol sets. Recurrent Neural Network-based algorithms may encounter difficulties with such expansive symbol sets. Additionally, some languages exhibit complexity in appearance and are sensitive to image attributes. Recognizing the quality of an integrated system for detecting and recognizing various languages, the paper emphasizes the need for an optimal solution capable of discovering compositional depictions that capture the diverse outlines of text cases across different languages.

In a related study, Amritha S Nadarajan et al. [7] focus on printed text detection within scene images. They employ essential image enhancement approaches for text detection, involving preprocess, text localization, classification of text, and text detection. Various classification techniques, including CNN, Adaboost, SVM, and Text-CNN, are employed. The comparative analysis presented in the paper concludes that CNN outperforms other techniques in this context.

X. Zhou et al. [8] presented an article, which describes and analyses the traditional earlier research works of text classification. In recent times, numerous research teams have contributed to automatic text detection, classification, and recognition. An extensive literature survey is available in this domain. Thus, interesting difficulties are still available in the domain. The research and development issues on how to make a breakthrough in the existing text categorization of huge text categorization problems, and how to shape much operative and effectual feature selection model have fascinated bags of research attention.

B. S. Anami and Deepa S. Garag [9] introduced a semi-automated approach to the cognizance of printed text in the Kannada language. This method incorporates human involvement to extract text, followed by automatic text recognition. The approach utilizes zone-based feature extraction to identify Kannada characters, and its performance has been authenticated through testing with various font styles. To evaluate performance, a three-fold

cross-validation method has been applied. The method shows promising results in recognizing printed primitives. Future research could extend the approach to recognizing handwritten kannada characters and assembling characters by connecting primitives at appropriate connection points. Such an extension could contribute to the digitization of old Kannada manuscripts, assisting archaeological efforts in reading Halegannada (old Kannada).

In the year 2020, authors proposed the scene text detection technique via a Simple least-square SVM approach [10]. This approach uses the Otsu threshold technique to set the threshold for the input scene images. A new methodology for scene text detection from the Multiview images is proposed [11]. This uses the Delaunay triangulation technique to detect the text from the scene images. The major drawback of this methodology is that limited only for two views.

In [12], a novel Multiscript-oriented scene text detection methodology was proposed by Roy S et. al., using the mutual nearest neighborhood concept. There is a restraint of being sensitive to poor-quality image inputs.

Guo J et al. introduced a cutting-edge technique for text identification aiming to identify traffic and text by exploring various color attributes extracted from images [13]. However, it is essential to note that this approach exhibits sensitivity to changing lighting conditions.

In 2020, a novel text detection method was proposed for recognizing text in native view images [14]. This method relies on morphological analysis of components for detection, but its performance is heavily dependent on the chosen sliding window sizes.

In 2019, an innovative approach for detecting scene text utilizing two Artificial Neural Networks (ANNs) was introduced to detect text from eight signboards [15]. However, this approach is limited in its ability to handle arbitrary orientations. Another approach for detecting text from both document images and natural scenes is proposed by identifying highly stable extremal areas. This method is particularly effective for low-quality images and those with low contrast.

In [16], a novel model to recognize Kannada text from the scene images are suggested using Semantic GAN and Balanced Attention Network (SGBANet). The Semantic GAN serves to extract the simple semantic features and Balanced Attention Module (BAM) recognizes the scene text. With the Semantic Generator Module (SGM) is operated to get the generation and discrimination on the semantic level. This generates the semantic simple features. The Semantic Discriminator Module (SDM) is used for differentiating the semantic characteristics from the scene text and clear text images. Also, the BAM improves the attention drift problem.

In citation [17], a system for detecting and segmenting arbitrarily oriented multi-language text is presented, employing a lineup of level sets and Gaussian Mixture Model (GMM). This method adopts a sequential feature processing approach, integrating a Gaussian low-pass filter and a single-level 2-D Discrete Wavelet Transform (DWT) to enhance feature extraction. The set of method used in addition to use the k-means algorithm and GMM, contributes to reliable character segmentation outputs. Recognition of multi-language text is facilitated by utilizing the Laplacian of Gaussian filter and morphological bridge function. The experimentation phase involves through the Multi-script Robust Reading Competition dataset and privately collected graphical and handwritten multi-language text images.The results demonstrate superior segmentation and detection capabilities compared to earlier techniques.

In [18], an improved text detection and recognition approach is implemented by the researcher to find the text detection results. Authors have accomplished the post-processing technique to improvise performance of the text detection and recognition process. This projected work obtains the key features, such as entropy, phase congruency, and compactness. For improving the feature extraction, authors have combined the SVM and Gaussian model for the text components. Resultant output provides the better results in detection of text. Authors have used MSRA-TD-500 and SVT dataset experimentation.

In 2020, this research article discovers the new approach which adapts for localizing and understanding the texts of natural scene images of Indian languages [19]. The deep learning technique is operated to detect the script representation of various Indian languages, despite this, the approach requires a significant improvement with the strong deep learning techniques.

In [20] details the deployment of a model for detecting and recognizing Telugu language scene text. This model is constructed using an end-to-end trainable neural network, specifically a CNN-Recurrent Neural Network-based architecture. Notably, the competence of this detection and recognition model can be widened to encompass other Indian languages.

In [21], The" Wiener Filter based Medical Image De-noising" introduces a technique using Wiener Filters to reduce noise in medical images, aiming to enrich their diagnostic accuracy and quality.

In [22], the paper "Character Recognition in Natural Images" delves into the intricate task of identifying characters within uncontrolled, real-world images. It likely explores a variety of techniques and models aimed at achieving robust character detection in diverse visual environments.

C. Yao, X. Bai, and W. Liu's [23], paper stands as a significant milestone in the realm of multioriented text detection and recognition. The unified framework they present not only addresses a critical limitation in existing OCR systems but also sets a benchmark for future research in this domain. The clarity in presentation, innovative methodology, and compelling experimental results collectively make this paper a must-read for researchers and practitioners in computer vision, image processing, and OCR.

M. Valdenegro-Toro, P. Plöger, S. Eickeler, and I. Konya's [24], paper makes a valuable contribution to the field of text detection and authentication in challenging environments, with a specific focus on road scenes. The innovative use of Histograms of Stroke Widths demonstrates a thoughtful approach to addressing the unique challenges posed by multi-script text in real-world scenarios. Researchers and professionals in computer vision, image processing, and intelligent transportation systems will find this paper to be a significant resource for advancing their understanding and exploration of text detection in complex environments.

M. R. Phangtriastu, J. Harefa, and D. F. Tanoto's [25],paper stands out as a valuable contribution to the field of OCR by offering a thorough analysis of the differences between Neural Networks and Support Vector Machines. The meticulous experimental design, clear presentation of results, and insightful discussions make this paper a valuable resource for researchers and practitioners in machine learning, pattern recognition, and OCR.

3. Methodology

In this proposed methodology, the authors introduce a pioneering approach. They leverage machine and deep learning techniques for the scene text detection and recognition process. The novel techniques incorporated in this proposed work boast unique features that prove instrumental in extracting features and recognizing kannada characters, words, or sentences.

This proposed work involves pre-processing, detecting, feature extraction and identification or recognition of kannada text are carried out after the collection of multilingual scene text images. Initially, the multilingual scene text images are collected and are pre-processed for effectual elimination of various noise and to refine the image quality. After pre-processing, detection step is performed for analyzing only kannada text by leaving the other texts. The highly discriminative text features will be pull out from detected kannada text. Finally, recognizing optimal Kannada language can be made in the arrangement of machine-readable strings. The proposed workflow of the methodology is showed in Fig.2. The novel methods utilized in this suggested project possess distinct attributes, which are very useful for extracting features and recognizing kannada characters or phrases or sentences.

The fundamental iteration of the algorithm is provided as below.

Begin algorithm

1. Input the multi-lingual scene text image.
2. Pre-process the input scene text image using bilateral and wiener filter.
3. Detect the kannada text from the pre-processed input scene text image using CNN-block attention-based YOLO v7 model.
4. Extract the key features using dense-stacked LSTM network model.
5. Recognize the kannada text using CNN-based residual network-assisted auto-encoder model.
6. Display the machine/human readable kannada text messages.

end algorithm

This proposed work involves pre-processing, detecting, feature extraction and identification or recognition of kannada text are carried out after the collection of multilingual scene text images. Initially, the multilingual scene text images are collected and are pre-processed for effectual elimination of various noise and for enhancing the quality of the image. After pre-processing, detection step is performed for analyzing only kannada text by leaving the other texts. The highly discriminative text features will be pull out from detected kannada text. Finally, recognizing optimal Kannada language can be made in the form of machine-readable strings.The proposed workflow of the methodology is demonstrated in figure 2.

Process flow of each step mentioned in the proposed model is described as follows.

3.1 Image Pre-processing

The various noises present in the multilingual scene text images are effectively eliminated using Twin chain filtering (TCF) technique. Here, the bilateral filter and wiener filter models are cascaded in series for enhancing the image quality and promote better detection.

3.1.1. Bilateral Filter

The Bilateral filter is a low pass filtering method to improve nonlinear and non-iterative data. The chief purpose of BF is smoothing of images with edge conservation. The process is achieved by replacing every pixel value by the average estimation. This filter depends upon the fundamental parameters where size, computational speed and contrast can be maximized over huge images. The elementary image corruption issues occur through existence of numerous noises. BFs are adopted to perform noise filtering specially in image processing and acts as a smoothing filter for noise

minimization. The bilateral filter is formulated as in the equation (1).

$$BF[I]_p = \frac{1}{Wp}\sum_{q\in S} G_{\sigma_s}(||p-q||)G_{\sigma_s}(|I_p - I_q|)I_{q_{0_0}} (1)$$

In equation (1), BF is the Bilateral filter, G is the normalized Gaussian function, s is the spatial extent, s is the minimum amplitude of the edge, Iq is the intensity at pixel q, The σ is the sigmoid in the color space. $[I]_p$ is the noisy image at pixel p, 1/wp is the normalization factor.

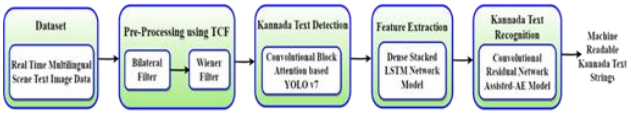

Fig. 2. Proposed Model in the Workflow: Streamlining Processes for Efficiency.

3.1.2. Wiener Filter

The Wiener filter (WF) supports minimizing the Root Mean Square Error that can manage both degradation functions and noises. The WF serves as one of the statistical filters that produces a suitable image estimation from a noisy time order. The kernel is tuned in WF to combine edge preservation and image details in accordance with efficient noise reduction obtained by executing a local Gaussian Markov random field. Here the images obtained from BF are fed into WF for the further eradication of noises. The effectively pre-processed images are fed into the detection model. The Wiener Filter in the Fourier Domain is represented as in equation (2).

$$G(u,v) = \frac{H*(u,v)P_S(u,v)}{|H(u,v)|^2 P_S(u,v)+P_n(u,v)} \quad (2)$$

Dividing through by Ps makes its behavior easier to explain as (3)

$$G(u,v) = \frac{H*(u,v)}{|H(u,v)|^2+\frac{P_n(u,v)}{P_S(u,v)}} \quad (3)$$

Here, H(u, v) represents the degradation function and H*(u, v) represents the degradation function's complex conjugate. Furthermore, Ps(u, v) denotes the power spectral density of the un-degraded image, while Pn(u, v) denotes the power spectral density of noise. One way to think of the ratio Pn/Ps is as the signal-to-noise ratio (SNR) inverse.

3.2 Detecting Kannada Text

From filtered multilingual scene images, the Kannada language is detected using Convolutional Block Attention based YOLO V7 model through bounding box generation. The YOLO series indicates to be a resulting detection technique depending on the deep machine learning and CNN. In this process, the main benefits include better detection rate, monitoring in real time and fast speed. YOLO V7 shows excellent performance in text detection that highly aims to enhance accuracy and detection speed. After the concatenation blocks of YOLO V7, the CBAM i.e., Convolutional Block Attention Model is added for considering most effective information. CBAM is an effective and simple module to the feed forward Convolution NN. For feature map, the CBAM infers the attentional map sequentially with two numbers of non-dependent dimensions which are known as space and

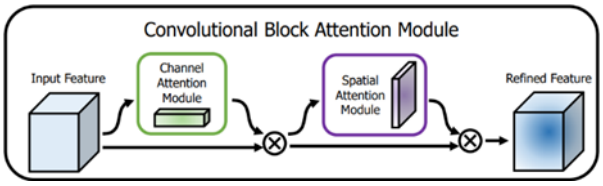

Fig. 3. Optimizing Layout: CBAM for Enhanced Performance and Efficiency.

channel. An attentional map gets multiplied by the input featured map then to effective weighted outcomes. The YOLO v7 model generates a bounding box for Kannada text by separating the images into several number of grids. The layout of CBAM is portrayed in Figure 3.

The CBAM progressively incorporates the two-dimensional spatial attention map $Nt \in S^{1*H*W}$ and the single-dimensional channel attention map $Ne \in S^{C*1*1}$, given an inter-mediate feature map $G \in S^{1*H*W}$ as an input. This is exemplified in Figure 3. The equations (4) and (5) provide outline of the entire attention process.

$$G' = Ns(G) \otimes G \quad (4)$$

$$G'' = Nt(G') \otimes G' \quad (5)$$

3.3 Feature extraction

Numerous attributes are obtained from the detected Kannada texts using Dense Stacked LSTM network model (D-SLSTM). Here, initially DenseNet-121 architecture is employed and the obtained traits from it are directed over the stacked LSTM model. The dense net concatenates the result of previous layer with the upcoming layer. The model has been intended particularly to enhance the degraded accuracy generated because of vanishing gradient issues. DenseNet-121 architecture is employed in this research work that applies the connections which are densed among the layers using Blocks of Dense by which each layer can be directly connected with one another. It's a feed-forward network that preserves some chief advantages like decreasing of vanishing-gradient issue, build up feature propagation, decreases

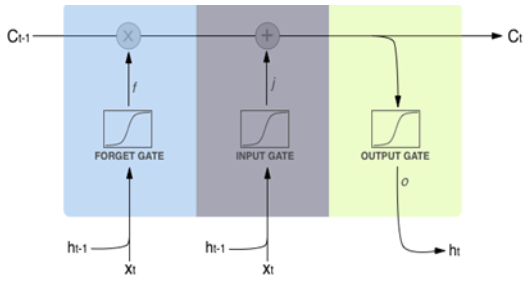

Fig. 4. Memory cell of LSTM

the parameter number and stimulate feature restoration. In stacked LSTM, two LSTMs are cascaded in series to obtain the most efficient features. The chief purpose of designing LSTM model is to resolve the back flow error problems. There are three gates presented in LSTM model including input, output and forget gate. Figure 4 shows the LSTM memory cell.

In Figure 4, Ct-1 and Ct are the cell state memories at timestamps t-1 and t, respectively.

The input gate, forget gate, and output gate equations of the LSTM memory cell are yielded by following equations (6) to (8), respectively.

$$i_t = \sigma(w_i + [h_{t-1}, x_t] + b_i) \quad (6)$$

$$f_t = \sigma(w_f + [h_{t-1}, x_t] + b_f) \quad (7)$$

$$o_t = \sigma(w_o + [h_{t-1}, x_t] + b_o) \quad (8)$$

In equations (6) through (8), it denotes the input gate, ft is forget gate and ot is the output gate. The σ is sigmoid function. The wi, wf, and wo are the weights of the respective gate neurons of the respective gates. xt is the present timestamp and ht-1 is the previous LSTM block result at timestamp t-1. At last, bi, bf, bo are the biasis for the respective gates.

The final resultant LSTM cell output $g_t^{(i)}$ is given in equation (9). This is attainable by closing the final output gate which uses a sigmoid unit for gating.

$$g_t^{(i)} = \sigma\left[b_i^0 + \sum_j u_{i,j}^0 x_j^{(t)} + \sum_j w_{i,j}^0 h_j^{(t-1)}\right] \quad (9)$$

where, b^0, u^0 and w^0 are biases, input and $x^{(t)}$ is the present timestamp for bias u^0 and similarly $h^{(t-1)}$ is the timestamp from the previous LSTM block.

3.4. Kannada Text recognition

Using extracted features, the optimal recognition of kannada language can be built in the form of machine-readable strings using Convolutional Residual network assisted auto encoder model (CResNet-AE). In this proposed work, authors have introduced CNN counterpart-based auto encoder into residual connection called CResNet-AE. This proposed framework is given in Figure 5 for instructing the model.

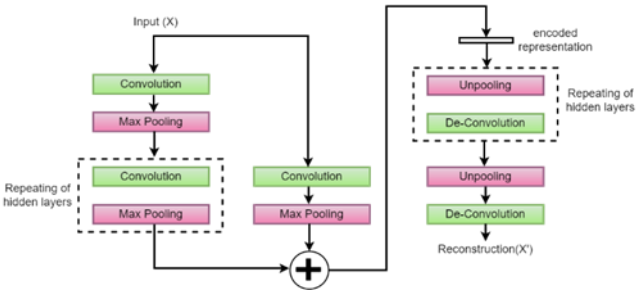

Fig. 5. Training the Model with CResNet-AE for Enhanced Performance and Accuracy.

The input, denoted by x and labelled as an n-dimensional vector corresponding to n neurons, is instructed on CResNet-AE to reconstruct the input. This process creates layers with the same number of neurons. The middle layer (not shown) consists of m neurons. Encoder E computes a hidden representation called z such that z=E(x). Decoder D then reconstructs the input and produces y=D(z). Both the encoder and decoder consist of multiple hidden layers forming a stacked RAE.

In the decoder, each hidden layer acts as a nonlinear unit, as shown in Equation (10).

$$\sigma(Vz + c) \quad (10)$$

where, σ is an activation function (sigmoid), V and c are the weight matrix and constant respectively. Authors use v^l to represent the weight corresponding to the layer l. In CResNet-AE, the product matrix is replaced with max-pooling and convolution operation as observed in Fig.5. Every hidden layer 'l' is collected by nonlinear mapping f (·) and a residual 'r'. Every hidden representation hl in hidden layer 'l' is calculated as in the equation (11).

$$h^{l+1} = r(h^l) + f(h^l) \quad (11)$$

The residual connection is represented as (12).

$$r(h) = w_r h \quad (12)$$

It serves as a linear function that validates the alignment of dimensions in the output of the 'f' function, representing a more compact network consisting of 'F' layers. Each layer within 'f' is a nonlinear mapping, consistent with the formulation in equation (13).

$$\sigma(Wh + b) \quad (13)$$

Where, W = weight matrix. The $W^{l,j}$ to represent the weight matrix which covers to the layer 'l' and sublayer 'j'. In

CResNet-AE W_r and W are switched by max pooling CNN operations.

Also, loss function J_θ of the CResNet-AE is also calculated by the difference of input(x) and output(y), which is given in (14)

$$J_\theta = \frac{1}{T}\sum_{i=1}^{T}\|x_i - y_i\|^2 \quad (14)$$

In this context, x_i represents the i^{th} input sample, y_i writes to the output for the i^{th} input sample, and θ encompasses the set of factors for the autoencoder, including weights and biases.

4. Results and discussion

For every research work, the results and discussions are the important stage. This proposed work is experimented with the more than 700 natural scene images.

Among these, authors have used 490 images (70%) to prepare the model and rest 210 (30%) to test the model i.e., recognition of Kannada character/words/sentence from natural scene images. In both the case i.e., training and testing, same

Fig. 6. Input and Output sample images.

methodology has been utilized. Figure 6 exhibits one of the sample input scene image considered for experimentation and its corresponding processed final outp ut image.

The assessment of the proposed work was conducted using five parameters in the following equations (15) through (19).

$$p = \frac{tp}{tp + fp} \quad (15)$$

In equation (15), p is the precision, tp and fp are the true positives and false positives respectively.

$$r = \frac{tp}{tp + fn} \quad (16)$$

In equation (16), r refers the recall, tp and fn are the true positives and false negatives respectively.

$$F = 2 * \frac{p*r}{p+r} \quad (17)$$

In equation (17), F is the F-measure or F-score, p and r are the precision and recall respectively.

$$MSE = \sum(y_i - p_i)^2 n \quad (18)$$

Within equation (18), MSE stands for Mean Square Error, where y_i denotes the ith observed value, and p_i symbolizes the corresponding predicted value for the i^{th} observation. The symbol Σ indicates that the summation is carried out over all the quantities of i.

$$A = \frac{n}{N} \quad (19)$$

A is accurate, wherein n represents the total of correct predictions, and N indicates the total number of samples utilized in the experimental study.

The proposed recognition system is compared among the following 5 existing systems as follows: Auto-Encoder, Resnet, Gated Recurrent Units (GRU), LSTM, Convolution Neural Network (CNN).

The graphical portrayal of the experimental outcomes are illustrated in the subsequent figures from Figure 7 to Figure 11 for F-Measure, MSE, Precision, Recall and Accuracy respectively. From the Figures 7, 9, 10, and 11 demonstrate superior F-measure, precision, recall, and accuracy in our proposed method compared to Auto-encoder, Resnet, GRU, LSTM, and CNN techniques.

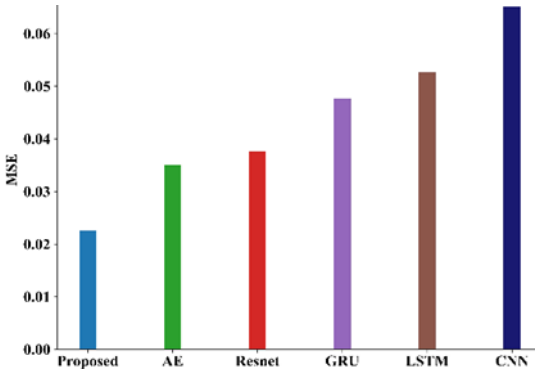

Fig. 7. F-Measure calculation.

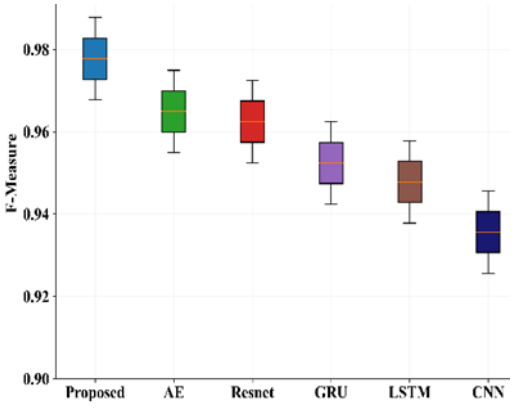

Fig. 8. MSE calculation.

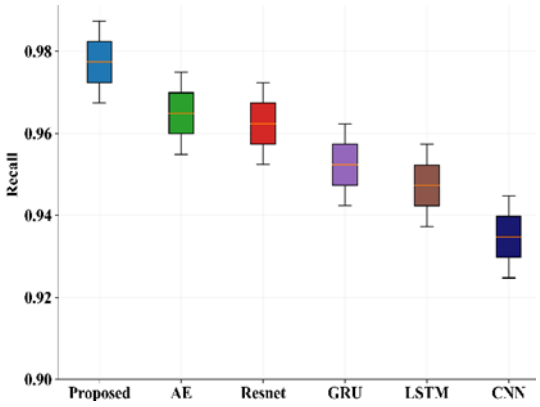

Fig. 9. Precision calculation.

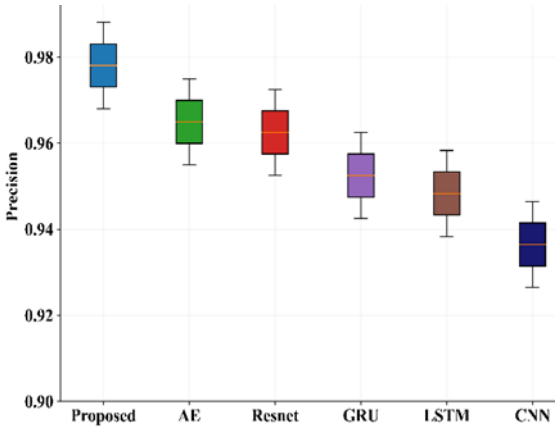

Fig. 10. Recall calculation.

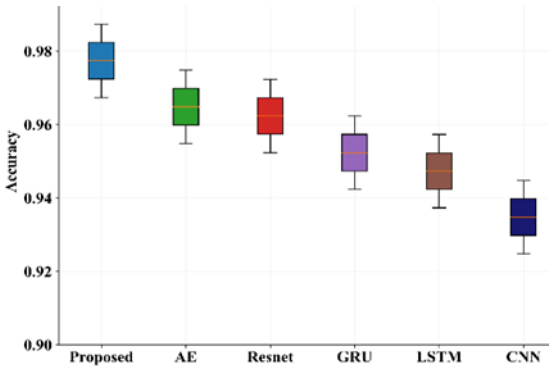

Fig. 11. Accuracy calculation.

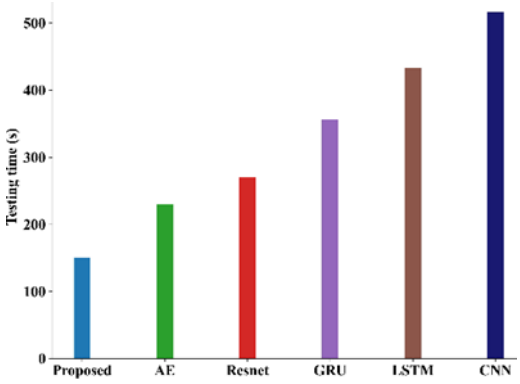

Fig. 12. Testing Time.

The system achieves a remarkable 97.8% recognition rate, surpassing other methods. Additionally, Figure 12 illustrates efficient testing time. The proposed system's MSE (Figure 8) outperforms existing techniques, notably CNN. Furthermore, Figure 12 highlights the significantly shorter testing time of our system (150 seconds) compared to CNN's 500 seconds, showcasing its efficiency.

5. Conclusion

With the advancement in artificial intelligence, In this research article, authors have attempted to recognize the kannada text from the natural scene images which contain images and other language text in it. Even, there are many research work carried out in scene text recognition, there is no or less research work carried out in kannada text recognition. This paper demonstrates the proposed work with five phases: reading the input, preprocessing, kannada text detection, various feature extraction, and recognition of kannada text (character/word/sentence) from the natural scene image. For preprocessing an input image, bilateral and wiener filters are used. The Convolutional Block Attention based YOLO-V7, and Dense Stacked LSTM Model are used for Kannada text detection and feature extraction respectively. In the last phase, Convolutional Residual Network assisted Auto Encoder model is used for recognition of kannada text. This paper highlights particular state of art techniques, proposed methodology flow diagram, and their brief discussion. In the fifth section of the paper, obtained results and their discussion has been carried along with the performance metrices and their mathematical equations and graphical representations of the results have been presented. It is observed that the proposed work has yields the better recognition rate of 97.8% and recognition time of 150 seconds which is more proficient than other techniques mentioned.

References

[1] Zheng Xiao et. al., An extended attention mechanism for scene text recognition. Expert Systems With Applications, 203, pp.1-11, May 2022.

[2] Sahana K Adyanthaya, Text Recognition from Images: A Study. International Journal of Engineering Research and Technology, Volume 8, Issue 13, pp.18-20, 2020.

[3] Matteo Brisinello et. al., Review on text detection methods on scene images. 61st International Symposium ELMAR-2019, Zadar, Croatia, pp.51-56, 2019.

[4] Xiaoqian Li,Jie Liu, Shuwu Zhang, Text Recognition in Natural Scenes: A Review. International Conference on Culture-oriented Science & Technology, pp.154-159, 2020.

[5] Vishnuvardhan and Dhanalakshmi Miryala. Scene

Text Recognition of Indian Languages in Natural Scene Images. International Journal of Advanced research in engineering and Technology [IJARET-2020].

[6] Shangbang Long, Xin He and Cong Yao, Scene Text Detection and Recognition: The Deep Learning Era. Received: 14 April 2020 / Accepted: 8 August 2020.

[7] Aggarwal, C. C. Neural Networks and Deep Learning: A Textbook. Basingstoke, England: Springer.2018.

[8] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, and S. Zhou, AON: Towards arbitrarily-oriented text recognition. In: Proceedings of CVPR, 2018, pp. 5571–5579, [CVPR-2018].

[9] Basavaraj S. Anami, Deepa S. Garag. A Semiautomatic Methodology for Recognition of Printed Kannada Character Primitives Useful in Character Construction. Recent Trends in Image Processing and Pattern Recognition. [Springer Singapore-2019].

[10] Francis, L. M., & Sreenath, N, TEDLESS: Text detection using least-square SVM for natural scene. Journal of King Saud University-Computer and information sciences, 32(3), 87–299.

[11] Roy, S., Shivakumara, P., Pal, U., Lu, T., & Kumar, G. H. (2020). Delaunay triangulation based text detection from multi-view images of natural scene. Pattern Recognition Letters, 129, 92–100.

[12] Raghunandan, K. S., Shivakumara, P., Roy, S., Kumar, G. H., Pal, U., & Lu, T, Multi-script oriented text detection and recognition in video / scene /born digital images. In:IEEE transactions on circuits and systems for video technology, pp. 1145–1161.

[13] Guo, J., You, R., & Hung, L, Mixed vertical and horizontal text traffic sign detection and recognition for street level scene. IEEE Access, 8, 69413–69425.

[14] Liu, S., Xian, Y., Li, H., & Yu, Z. (2020). Detection in natural scene images using morphological component analysis and Laplacian dictionary. IEEE Journal of Automatic Sinica, 7(1), 214–222.

[15] Panwar, M. A., Memon, K. A., Abro, A., Zhongliang, D., Khuhro, S. A., & Memon, S. (2020). Signboard detection and recognition using artificial neural networks. In: Proceedings on ICEIEC, pp. 16–19.

[16] Dajian Zhong et. al., SGBANet: Semantic GAN and Balanced Attention Network for Arbitrarily Oriented Scene Text Recognition. arXiv:2207.10256v1 [cs.CV], July 2022.

[17] H. T. Basavaraju et. al., Arbitrary oriented multilingual text detection and segmentation using level set and Gaussian mixture model. Evolutionary Intelligence, Vol. 14, pp.881– 894, 2021.

[18] Hamam Mokayed et. al., A new defect detection method for improving text detection and Recognition performances in natural scene images. IEEE Explorer.

[19] A. Vishnuvardhan, and Dhanalakshmi Miryala, Scene Text Recognition Of Indian Languages In Natural Scene Images. International Journal of Advanced Research in Engineering and Technology, Vol. 11, Issue.12, pp. 2773-2781 , 2020

[20] A. Ram Bharadwaj et. al.,Telugu text extraction and recognition using convolutional and recurrent neural networks. International Journal of Engineering and Advanced Technology, pp.1449-1451, Vol.8, Issue-5, 2019.

[21] Dr. Sana'a khudayer Jadwa,Wiener Filter based Medical Image De-noising. International Journal of Science and Engineering Applications Vol. 7–Issue 09, pp.318-323, 2018

[22] T. E. De Campos, B. R. Babu and M. Varma, Character recognition in natural images. VISAPP 2009 (2).

[23] C. Yao, X. Bai and W. Liu, ”A Unified Framework for Multioriented Text Detection and Recognition. In:IEEE Transactions on Image Processing, vol. 23, no. 11, Nov. 2014, pp. 4737-4749.

[24] M. Valdenegro-Toro, P. Plo¨ger, S. Eickeler and I. Konya,Histograms of Stroke Widths for Multi-script Text Detection and Verification in Road Scenes.2016 IFAC-Papers Online, 49(15), pp. 100–107. [Online].

[25] M. R. Phangtriastu, J. Harefa, and D. F. Tanoto, Comparison between Neural Netwok and Support Vector Machine in Optical Character Recognition. Procedia Comput. Sci., vol. 116, 2017, pp. 351–357.

[26] Janani, S., Dilip, R., Talukdar, S.B., Talukdar, V.B., Mishra, K.N., Dhabliya, D. IoT and machine learning in smart city healthcare systems (2023) Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, pp. 262-279.

[27] Juneja, V., Singh, S., Jain, V., Pandey, K.K., Dhabliya, D., Gupta, A., Pandey, D. Optimization-based data science for an IoT service applicable in smart cities (2023) Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, pp. 300-321.

.