IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 March 2023, accepted 31 March 2023, date of publication 5 April 2023, date of current version 2 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264569

RESEARCH ARTICLE

Enhanced Feature Model Based Hybrid Neural Network for Text Detection on Signboard, Billboard and News Tickers

S. ANBUKKARASI[1], VEERAPPAMPALAYAM EASWARAMOORTHY SATHISHKUMAR[2], C. R. DHIVYAA[1], AND JAEHYUK CHO[2]

[1]Department of Computer Science and Engineering, Kongu Engineering College, Erode, Tamil Nadu 638060, India
[2]Department of Software Engineering, Jeonbuk National University, Jeonju-si, Jeollabuk-do 54896, Republic of Korea

Corresponding author: Jaehyuk Cho (chojh@jbnu.ac.kr)

This work was supported in part by the Korea Environmental Industry and Technology Institute (KEITI) funded by the Korean Government, Ministry of Environment (The Development of IoT-Based Technology for Collecting and Managing Big Data on Environmental Hazards and Health Effects), under Grant RE202101551; in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) funded by the Korean Government, Ministry of Science and ICT (MSIT) (Implementation of Verification Platform for ICT Based Environmental Monitoring Sensor), under Grant 2019-0-00135; and in part by IITP funded by the Korean Government, MSIT (Building a Digital Open Laboratory as Open Innovation Platform), under Grant 2021-0-00546.

ABSTRACT Recognizing text from the nature scene images and videos has been the challenging task of computer vision and machine learning research community in recent years. These texts are difficult to recognize because of their shapes, complex backgrounds, color, shape and size variations. However, text recognition is very much useful in indexing, keyword-based image and video search, and information retrieval. In this research paper, a model is proposed to detect the isolated text characters in the photographic images of natural scenes. The proposed model uses the combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) for recognizing the text in natural images. The model uses two networks, where the first network combines the low-level and middle-level features to increase the feature size and passes the enriched information to the second network. Here, features are again widened by combining with high-level features, resulting in powerful and robust features. To evaluate the proposed model, ICDAR2003 (IC03), ICDAR2013 (IC13), SVT (Street View Text) datasets have been used. And an extensive Tamil news tickers image dataset has been developed to evaluate the model. The experimental results show that the combined feature fusion technique outperforms the other methods on the ICDAR2003, ICDAR2013, SVT and Tamil news tickers datasets.

INDEX TERMS Natural scene recognition, machine learning, deep learning, image processing, natural scene Tamil character recognition.

I. INTRODUCTION

The rapid growth in technologies allows us to acquire a huge number of images and obtain a variety of natural scene pictures in real time every day. The natural scene picture includes several outdoor environmental images such as billboard images, sign board images, images on products, news tickers, and street signs. In this research paper, the proposed work focuses on billboard images, sign board images, and Tamil news tickers to recognize the text on images. The billboard images are used to advertise the product's brand to a variety of customers. It contains information about products, such as the website of the product and a phone number to provide additional details. For instance, new pipeline is proposed in [8] to recognize the website of the products from the billboard, and automatically the official website is connected for delivering more information about the products to the customers. The sign board images are designed for identifying

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

the particular organization, finding the exact location, providing safety information and marketing the product. A neural network [6] is enhanced to observe the signboard's edges for detecting the text on signboards from the environment. The proposed work uses the Street View Text (SVT) dataset and the ICDAR2003 dataset to identify the English text on signboard images as well as on bill board images.

News channels are required to transmit the information across the world, and news tickers display the latest news in a line of text that moves across the bottom of television screen. The majority of the work for detecting text on news tickers images has been done in the literature, specifically for English text [4], Chinese text [12], Urdu-text [2] and Arabic text languages. Although some research has been proposed to recognize hand written Tamil language [10], there is no well-known research study to recognize the Tamil text on news tickers images. Accordingly the proposed research work mainly concentrates on detecting the Tamil text which helps the news analyst to generate the text report for their analysis. It reduces manual work, time, and self-effort, too. In this paper, Tamil news tickers are collected from around 12 Tamil news channels and new dataset is created for detecting Tamil text on news tickers effectively.

The images contain visual information and also numerous pieces of textual information, which is essential for many real time applications such as navigating autonomous vehicle, retrieving images based on contents, assist the visually impaired people and translating languages for tourists. Hence, recognizing the text from the image is crucial for text based applications. The traditional text recognition techniques have been developed to find the text from the images, but the performance of detecting the text is limited due to the variations in colors, font sizes, styles, lights, orientations, and the poor background of the image.

The unavailability of standard Tamil news tickers datasets is another limitation to evaluating the existing techniques for addressing the character recognition problem. A huge dataset is primary requirement to train and test the classifiers for Tamil text recognition. There is no standard dataset available for Tamil news tickers to analyze the Tamil text detection systems. Therefore, a new dataset is created in this research study. The dataset contains images of Tamil news tickers containing Tamil text which are collected from Tamil news channels. After performing the preprocessing operations, the uniform and standard dataset is passed to the classifier for recognizing the Tamil text from the images. The accurate classification of Tamil characters from the image of the Tamil news ticker is a very challenging task. Due to this, Researchers have not analyzed the Tamil languages in depth and this research work is still its in beginning stages.

To attain the best Tamil text recognition accuracy, it is required to build a powerful feature extraction model that can determine the significant features from the input image. However, in this research work, an improved CNN with Manifold feature extraction neural network and Multi-layer feature aggregation neural network [52] are proposed to extract significant features such as low-level, high-level features and aggregate them [29] [31] to identify the Tamil characters in Tamil news ticker images. When a CNN is trained on a set of images, the convolutional layers learn to recognize patterns in the images that are relevant to the task at hand. Once the CNN has been trained, it can be used to extract features from new images. This is typically done by passing the image through the network and extracting the output from one of the intermediate layers. The output from each layer corresponds to a different set of learned features, with earlier layers capturing low-level features such as edges and textures, and later layers capturing higher-level features such as object parts and scenes. These extracted features can then be used as input to the Recurrent Neural Network bidirectional Long Short Term Memory (RNN-BiLSTM) model [16], [30] which is used for recognizing frame sequence, and finally the decoder layer is used for decoding to predict the characters in an image by translating the frame sequence into end result. The performance of proposed research work is compared with sequential Convolutional Neural Network (CNN) [15], RNN-LSTM (Recurrent Neural Network Long Short Term Memory) and RNN-GRU (Recurrent Neural Network Gated Recurrent Unit) models. The experimental results of our proposed model provide superior results, and it outperforms the other three models. In addition, our proposed research model is evaluated on the SVT dataset, ICDAR2003 dataset and ICDAR2013 dataset where the proposed technique achieves better accuracy on the SVT dataset and also attains competitive accuracy results on the ICDAR2003 and ICDAR2013 datasets. The objective of this research is given below.

- To improve the performance of the model, improved CNN with Manifold feature extraction neural network and Multi-layer feature aggregation neural network are proposed by considering all the levels of the features.
- The deep learning techniques such as CNN, RNN LSTM and GRU models with the proposed feature extraction technique are used for detecting the text on the image.
- Various images that bear Tamil characters are collected and preprocessed for contributing to the research community of one of the low resource regional languages called Tamil.

The major contributions of this paper are:

- The dataset has been built with the images that consist of various Tamil characters, as Tamil is one of the languages with the most alphabets.
- Presenting new feature extraction methods that deal with various feature levels of the data.
- Comparing the proposed model with various existing techniques and evaluating it with various datasets.

The rest of our research paper is organized as follows: Section II describes related work. Section III presents our proposed methodology which explains the feature extraction techniques, the sequence recognition model and the decode layer. Section IV elaborates on the implementation details and experimental results. Finally, Section V illustrates the conclusion and future work.

II. RELATED WORK

Identification of text from image has been done using various methodologies, such as statistical methods and neural network methods. Earlier works deal with hand written character recognition as well as Optical Character recognition (OCR). Recognition of characters on sign boards, billboards and news stickers are getting attention in recent days. For handwritten Tamil characters, neural network based classification is performed in [5]. Artificial Feed Forward Network (ANN) is used as the classification model to identify the scanned characters in a given image. The hand written characters are collected and scanned to create the dataset. The Zoning method is used to segregate the images for extracting the features of the characters. Authors haven't given the accuracy details of their model in the paper. For Tamil vowel characters recognition, deep learning methodology is used in [11]. After the image pre-processing step, data augmentation is carried out. Image shearing, rotation and scaling have been performed as part of the data augmentation technique [50]. A Deep Belief Neural Network is implemented in Theano for classifying the given input characters. The authors claimed that they achieve 99% accuracy for their dataset.

For English alphabet recognition, neural network approach without extracting features is used in [7]. The input dataset contains 4840 characters. The given input images are resized and segmented for training the model. The algorithm performed internal segmentation as well as external segmentation on the given input images. Hidden layers with units 70, 40 and 30 are used. Authors claimed that the model produces 91% accuracy for the give data set. Arabic text recognition from news videos is performed and new dataset has been created in [13]. The dataset they created includes eighty videos, which consist of 850,000 frames. These videos are collected from four Arabic news channels. A region based text detection method is used for character recognition. Performance of the system is measured using Precision, Recall, and F-measure evaluation metrics. They claim that their system yields 70% of F-measure.

Edge based localization of text regions in video is given priority for text materials in [14]. They proposed a hybrid approach for English ticker identification from news videos. For the low quality videos, some enhancement operations on contrast are applied for enriching the video frames, and ticker text regions are segmented using morphological operators. Authors claim that their model produces good results on good quality videos as well as on bad quality videos.

A system is developed to identify the object and text in the given image in [8]. They developed a GUI to get the image as input and the system takes the user to the product website. They have used four different datasets for training the model. For object detection, Convolutional Neural Network (CNN) is used for text detection and recognition, a network called Tessaract OCR is used. The authors claim that their system outperforms Google image search baseline by saying that searched websites always appear in the top ten in the final rank. Reference [9] presents the first comprehensive dataset for Urdu language to recognize the Urdu news ticker text. They used multi-faceted OCR framework for recognizing the text in news ticker videos. Bi-directional Long Short Term Memory (Bi-LSTM) is used for news ticker text recognition.

A framework is created to recognize and detect text in a natural environment [6]. Texts are captured from signboards using smart devices, and edges are detected. The detected text is applied with algorithms and identified in two different languages such as Urdu and English. In phase1, an Artificial Neural Network (ANN) is used to detect the text from the signboards. Authors claim that 70% accuracy is achieved in text recognition on sign boards. Multilevel Convolutional Neural Network fusion technique is used in [1] to recognize the cursive Urdu texts in natural scene pictures. They implemented Multi Scale Feature Aggregation (MSFA) and Multilevel Feature Fusion (MLFF) [53] for recognizing the isolated Urdu characters. It performs the aggregation of multiple scale features [51] and merges these features with high-level features. The result of the networks is combined together to create the strong features. With this combined feature model, they achieved a 91% F-score when compare to a single line model. Dynamic Programming with two interconnected ANNs [3] is used for text line recognition. The character recognition is performed from the line segment with language independent features in a synthetic dataset. Images with low quality, complex background, different types of language, and fonts are considered in their work. From medical laboratory reports, text is detected and recognized in [48] using a deep learning approach. A patch based strategy is used for text detection in the documental images. The authors claimed that they achieved an Average Precision (AP) of 90%. An attention based scheme is used to detect scene text in [37]. They implemented multi oriented text detection with attention mechanism. It helps to detect the text on pixel by pixel basis. For the regression loss function, the diagonal adjustment factor is included. The author claimed that this factor increases the F-score by 0.8.

From the literature survey, it is evident that text detection and recognition is an emerging research area. The detection of the text for such low resource language is not yet implemented but it plays crucial role in image processing and Natural Language Processing domain. From the above discussion, it is known that these kinds of works are lacking in feature extraction part to improve the system performance. No work is reported for improving the feature extraction mechanism which plays an important role in text detection and recognition. Hence this work focuses more on feature extraction model with significant improvement in performance.

III. PROPOSED METHODOLOGY

In the proposed work, the given sets of images are preprocessed and features are extracted using MFEN and

MLAN techniques. Deep learning based RNN model is used for recognizing the text label sequences from the input images.

The decoder layer is used to translate the frame sequence for predicting the text on the image. All these steps are described in detail below. An improved CNN architecture is proposed to manage the problem of Tamil text detection in Tamil news tickers images. The proposed architecture integrates the features of different convolutional layers in the neural network and combines these features with high-level layer features to generate aggregated features. The extracted features are fed into the bidirectional Long Short Term Memory (Bi-LSTM) networks to detect the frame sequences from the input images. Finally, a decoder layer is applied to predict the Tamil text on input images. In further experimental analysis, sequential Convolutional Neural Network, RNN-LSTM (Recurrent Neural Network Long Short Term Memory) and RNN-GRU (Recurrent Neural Network Gated Recurrent Unit) models are developed to identify the Tamil characters from the Tamil news tickers images and their Tamil text recognition performance is compared with our proposed model. The proposed model consists of preprocessing the input image, Baseline network, feature extraction techniques, a Bi-LSTM for frame sequence detection and decoder layer. The architecture of proposed network is shown in Figure 1.

A. PRE PROCESSING

In our research work, initially the input image is resized to a fixed size. The fixed size contains 48 pixels of width, 48 pixels of height, and 3 channels. It is represented by its size of 48 × 48 × 3. The resized image is converted into a gray scale image, which is represented by 48 × 48 × 1, and it is normalized before it is passed to the next step in feature detection. For the purpose of improving the accuracy of text recognition, Gaussian filters [1] are applied to remove the noise from the input images which helps CNN to identify more significant textual features effectively.

B. BASE NETWORK

The baseline network is equipped with two convolutional layers. For the convolutional layers, the numbers of filters used are 64,128,256,512 and the size of the filter is 3. The dimension of the output feature vector is maintained by padding the values in image border. This layer is followed by a ReLU activation layer for non-linearity. A stride value of [2, 2] and Kernel size of [2, 2] at the max pooling layer was applied after all the convolutional layer. The size of the Kernel filter is varied between the values of 22, 33, and 55, and the promising accuracy was achieved with a kernel size of 55. Usually, in scenery images, text is placed randomly and its size is remarkably small when compared with the dimension of the whole image. For this reason, the filter size is kept small to take out more information from Tamil and English character images.

C. FEATURE EXTRACTION

The proposed model uses two different components for feature handling. 1. Manifold Feature Extraction Neural network (MFEN) and 2. Multi-layer Feature Aggregation Neural network (MLAN).

1) MANIFOLD FEATURE EXTRACTION NEURAL NETWORK (MFEN)

This component aggregates the various features of convolutional layer at diverse levels. The network complexity is reduced in CNN, by sharing the weight across different spatial areas by representing the features. Features from images are aggregated by gathering various hierarchical features. In this network, convolutional layers with 1 x 1 kernel are used, which in turn reduce the feature map dimensions and 256 output channels [1]. With the help of sampling and summation, middle level and low level features are combined with different channels. It uses the nearest neighbor interpolation technique to widen the feature map with lesser dimension of size same as the larger ones. This combined features maps gives enriched information when compare to individual features. Features are gathered at multiple scales to expand the size of feature map.

2) MULTI-LAYER FEATURE AGGREGATION NEURAL NETWORK (MLAN)

This network part combines the features from MFEN component with high-level features. The input is passed as aggregated features from the MFEN network to MLAN network. This network is built with two fully connected layers with 256, 512 output units. This network combines the high-level features with the features obtained from MFEN and passes the output to the RNN. This sequential RNN consists of Bi-LSTM model that processes the single feature vector given by the MLAN network. The feature map from MLAN is represented as,

$$\text{MLAN} = \text{Aggregate}\ (\text{MFEN}, \text{High level features}) \quad (1)$$

By combining the features, robust information is obtained from the images to recognize the characters effectively.

D. RECURRENT NEURAL NETWORK FOR SEQUENCE IDENTIFICATION

Traditional neural networks lack the contextual based information from the previous learning. Recurrent neural networks solve this issue by allowing information to be passed from one time step to the next. It has the ability to hold the information from the previous steps. Vanishing Gradient problem in simple RNN is solved by using a special kind of RNN called Long Short Term Memory (LSTM) [18]. A LSTM cell consists of three gates. 1. Input gate (i); 2. Output gate (o); and 3. Forget gate (f). It has the responsibility to judge which details must be hold or removed from the cell state.

The sigmoid layer in forget state determines whether the details should be regained in the cell state, as given by the

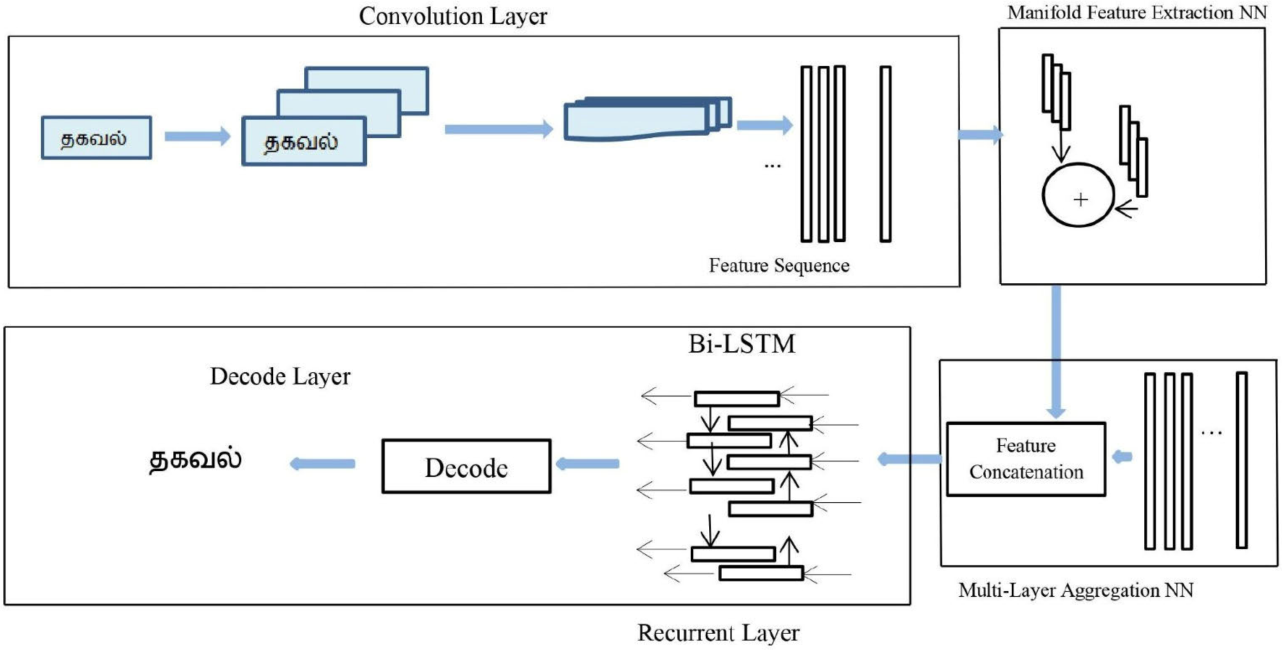

FIGURE 1. Architecture of MFEN-MLAN with Bi-LSTM model. The given input is passed to various layers and the features are extracted as sequence. These features are extracted and concatenated. The combined features are passed to Bi-LSTM layer and decoded.

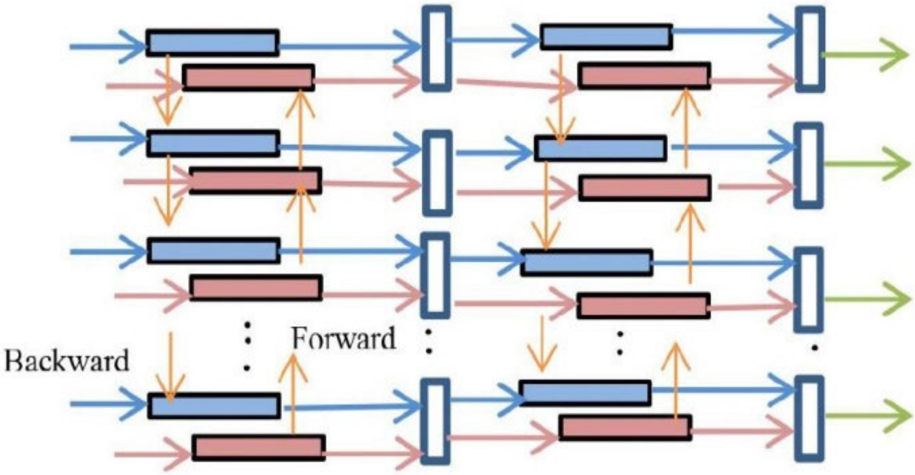

FIGURE 2. Structure of Bi Long Short-Term Memory Neural Networks, the red color blocks indicates the forward layer and blue color blocks indicates the backward layer.

following equation.

$$g(x) = \frac{1}{1+e^{-x}} \tag{2}$$

It can be even rewritten as,

$$g_j = \sigma\left(W_j\left\{h_{t-1}, x_t\right\} + b_j\right), \quad j\varepsilon(f, i, o) \tag{3}$$

W{h,t} $\longrightarrow$ Weight matrix, b_j $\longrightarrow$ bias vector

h_{t-1} represents the hidden state at the time stamp t-1, x_t represents input sequence at the time stamp t. The output of the sigmoid layer is g_j, the value between 0 or 1. So it can be represented as,

$$g_f = \sigma\left(W_f\left\{h_{t-1}, x_t\right\} + b_f\right) \tag{4}$$

The input gate considers the information of which new details should be hold in the cell state. It will be decided by new candidate values passed. The input gate is represented by the following equation.

$$g_i = \sigma\left(W_i\left\{h_{t-1}, x_t\right\} + b_i\right) \tag{5}$$

New candidate value can be identified by,

$$C_{vt} = tanh\left(W_{Cv}\left\{h_{t-1}, x_t\right\} + b_{Cv}\right) \tag{6}$$

The two statements are used to generate new value C_t. LSTM updates the information based on,

$$C_t = C_{t-1}.g_f + g_i.C_{vt} \tag{7}$$

In the proposed work, two bidirectional LSTM networks are used to calculate the hidden state. Bi-LSTM networks are very useful to learn long term dependencies. Figure 2 represents the Bi-LSTM with forward and backward directions.

E. DECODER LAYER

The decoder layer detects the highest probability of each frame to predict the correct label sequence. In our work, the decoding task is performed by non-lexicon based mode in which the predictions are performed without using any lexicon and our work is also compared with lexicon based sequence recognition. In this work, Connectionist Temporal Classification (CTC) [17] is considered as loss function which calculates the loss to update the parameters.

The decoding task is applied to predict the label sequence M∗ from the given input sequences and it is defined by

$$\boldsymbol{M* = arg\ max_a\ p\left(a\,|\,y\right)} \tag{8}$$

$$\boldsymbol{p\left(a\,|\,y\right) = \sum\nolimits_{\beta:\varphi(M*)=N} p\left(\beta\,|\,y\right)} \tag{9}$$

where $\beta = \{\boldsymbol{\beta_T}\}$, T€[1,t] is described as a sequence label from the time 1 to time t and the probability of β is calculated by

$$\boldsymbol{p\left(\beta\,|\,y\right) = \prod\nolimits_{T=1}^{t} y_{\beta_T}^{T}} \tag{10}$$

In non-lexicon based mode, the maximum value is chosen to find the most feasible label and the text label can be produced by

$$\boldsymbol{N = \varphi(arg\ max_a\ p\left(a\,|\,y\right)).} \tag{11}$$

IV. EXPERIMENTS

The proposed network model is evaluated on two English character datasets that are publicly available and our own Tamil news tickers image dataset. This Tamil news ticker image dataset can be considered as gold standard dataset, which is one of our contributions to this work. The performance of text recognition on image is compared with other well-known models, and the results of proposed work are analyzed with a number of state-of-the-art text detection techniques

A. DATASETS

In our proposed work, the datasets, specifically ICDAR2003 (IC03), ICDAR2013 (IC13), SVT (Street View Text) datasets and Tamil news tickers image dataset that we created, are used for evaluating the performance of the text recognition on images.

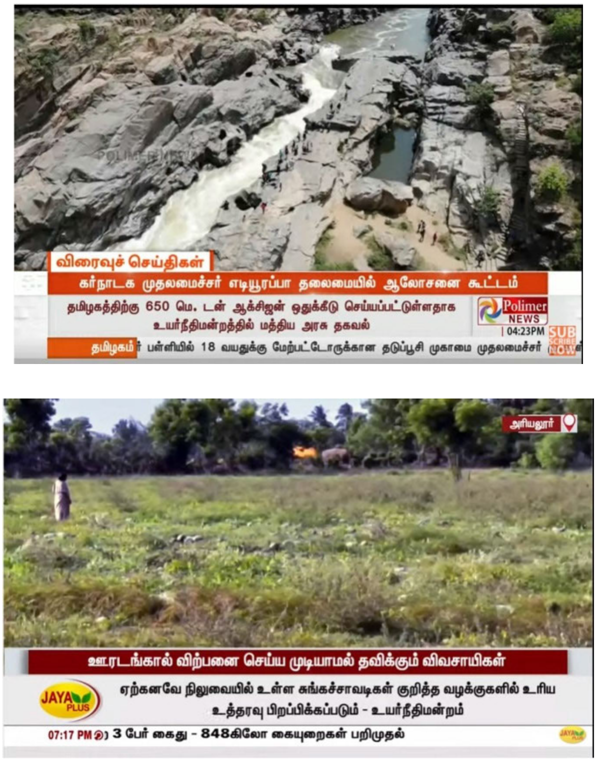
விரைவுச் செய்திகள்
கர்நாடக முதலமைச்சர் எடியூரப்பா தலைமையில் ஆலோசனை கூட்டம்
தமிழகத்திற்கு 650 மெ. டன் ஆக்சிஜன் ஒதுக்கீடு செய்யப்பட்டுள்ளதாக உயர்நீதிமன்றத்தில் மத்திய அரசு தகவல்
Polimer NEWS
04:23PM
SUBSCRIBE NOW
தமிழகம்
ர் பள்ளியில் 18 வயதுக்கு மேற்பட்டோருக்கான தடுப்பூசி முகாமை முதலமைச்சர்
அரியலூர்
ஊரடங்கால் விற்பனை செய்ய முடியாமல் தவிக்கும் விவசாயிகள்
JAYA PLUS
ஏற்கனவே நிலுவையில் உள்ள சுங்கச்சாவடிகள் குறித்த வழக்குகளில் உரிய உத்தரவு பிறப்பிக்கப்படும் - உயர்நீதிமன்றம்
07:17 PM
3 பேர் கைது - 848கிலோ கையுறைகள் பறிமுதல்

FIGURE 3. Sample Images of Tamil news tickers image dataset.

Tamil news tickers image dataset contains 31 basic Tamil characters which have 12 vowels characters, 18 Tamil consonants and one special character namely the āytam in their various forms. The 216 compound characters are formed by combining the vowels and consonants, so there are a total of 247 characters in Tamil language. We captured 300 total images of Tamil news tickers images from around 12 Tamil news channels, of which 125 images were used for training and 175 images were used to test the performance of our proposed model. The sample Tamil news ticker images of our dataset are shown in Figure 3.

ICDAR2003 (IC03) [19] dataset contains signboard images, bill board images, and also natural scene images for recognizing the text on images. In this dataset, the total number of images is 507, which includes 258 images to train the system and 249 images to test the text recognition system. The 860 cropped images of this dataset can be used to analyze the results of the proposed network. It includes alpha numeric characters and each word has at the minimum of 3 characters.

ICDAR2013 (IC13) dataset contains a test set of 1015 text images, most of which are derived from the ICDAR2003 dataset

Street View Text (SVT) [20] dataset contains 350 street view images, which include 101 images to train the model and 249 images to test the character detection model. These images were collected from Google street view, and 647 text images are cropped for analyzing the performance of our research work. Each text image has fifty word lexicons, described by Wang et al. [20].

TABLE 1. Hyper parameters of the proposed model.

Parameters	Values
Loss Function	sparse categorical cross-entropy
Optimizer	stochastic gradient descent
Activation	Relu
Batch Size	64
Epoch	80
Learning rate	0.01

B. IMPLEMENTATION DETAILS

The proposed work is implemented using a Python-based deep learning open source library called Keras [23], which runs on machine learning platform called TensorFlow. Images are shuffled and randomly segregated into training (80%) and testing (20%) datasets. Images are resized to 40*4*3 pixels. For Tamil news tickers dataset, we used data augmentation technique of the Keras library to boost the accuracy of the model. The training image samples are augmented for data expansion. Data augmentation is not performed on English image dataset. We compared the results of the proposed method with and without augmentation technique.

Random search for tuning the hyper parameters of the model

Random search hyper parameter tuning is applied to optimize the performance of a model by selecting the best hyper parameters from a defined search space. Hyper parameters can have a significant impact on the performance of a model. In the random search hyper parameter tuning, a random selection of hyper parameters is chosen from the search space, and the corresponding model is trained and evaluated on a validation dataset. This process is repeated multiple times with different randomly chosen hyper parameters until the best set of hyper parameters that results in the highest performance is found.

The best hyper parameters of the model found by random search are depicted in Table 1. The Sparse categorical cross-entropy loss function [21] with back-propagation and stochastic gradient descent optimization with momentum [22] was used for training the model. Batch size with 64 gave promising result. Network is trained with 0.01 learning rate and 80 epochs. The network contains around 10 million parameters.

C. PERFORMANCE METRICS

The performance of our proposed model is calculated and analyzed using standard valuation metrics known as precision and recall. These metrics are used to measure the accuracy of the text recognition system. Precision computes the number of true text class predictions called positive class and it is determined by

$$\mathit{Precision} = \frac{\mathit{True\ Positive}}{\mathit{True\ Positive} + \mathit{False\ Positive}} \tag{12}$$

Recall computes the number of true text class predictions made out of positive sample images in the dataset and it is evaluated by

$$\mathit{Recall} = \frac{\mathit{True\ Positive}}{\mathit{True\ Positive} + \mathit{False\ Negative}} \tag{13}$$

After calculating precision value and recall value, the overall classification accuracy of the proposed model is measured using F-score value and it is defined by

$$F - \mathit{Score} = \frac{2 * \ \mathit{precision} * \mathit{recall}}{\mathit{precision} + \mathit{recall}} \tag{14}$$

TABLE 2. Performance of our research model on the Tamil news tickers images dataset and comparison with other models.

Models	Precision (%)	Recall (%)	F-score (%)
CNN	80	81	81
RNN-LSTM	83	84	84
RNN-GRU	85	86	86
RNN-Bi-LSTM	87	88	88
MFEN-MLAN with Bi-LSTM	90	91	91

D. CLASSIFICATION RESULTS OF MFEN-MLAN WITH BI-LSTM MODEL

Classification results for Tamil characters in Tamil news ticker images with our proposed research work is compared with other well-known models such as sequential Convolutional Neural Network(CNN) [15], RNN-LSTM (Recurrent Neural Network Long Short Term Memory), RNN-GRU (Recurrent Neural Network Gated Recurrent Unit) models and RNN-Bi-LSTM(Recurrent Neural Network Bidirectional Long Short Term Memory). The performance metrics are calculated for our proposed model and the comparison results are shown in Table 2 and Figure 4. As comparison made in Table 2, the proposed feature aggregated method with RNN-BiLSTM network model achieves 90% precision, 91% recall and 91% F-score to recognize the Tamil characters on Tamil news ticker images whereas other three models provide less F-score value than the proposed model.

The precision, recall and F-score value for the CNN model is 80%, 81% and 81% respectively and these scores for RNN-LSTM, RNN-GRU methods are 83%, 84%,84% and 85%,86%,86% respectively. The RNN-Bi-LSTM achieves better results with 87% precision, 88% Recall and 88% F-score. Hence, Bi-LSTM is integrated with our proposed model MFEN-MLAN to achieve the best results than other models.

Due to the limited size of our Tamil news tickers image dataset, the data augmentation technique was applied to create the virtual images by scaling, cropping and translating the original Tamil news tickers images. With augmented data, the proposed model was trained to recognize the Tamil text on news ticker images and the results are compared with other models as summarized in Table 3 and Figure 5.

The proposed model with data augmentation upgrades the accuracy of Tamil text recognition by 3% and comparatively it achieves better result than non-data augmentation method.

TABLE 3. Performance of our research model on the tamil news tickers images dataset with data augmentation and compare the results with other models.

Models	Precision (%)	Recall (%)	F-score(%)
CNN	81	82	82
RNN-LSTM	84	85	85
RNN-GRU	86	87	87
RNN-Bi-LSTM	89	90	90
MFEN-MLAN with Bi-LSTM	93	94	94

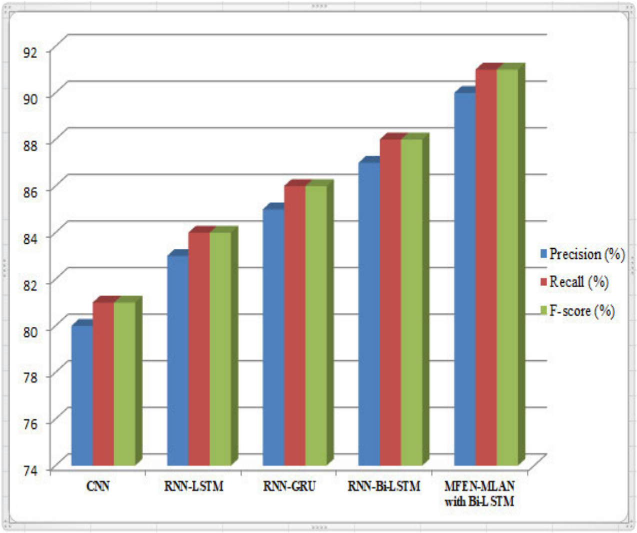

FIGURE 4. Comparison of proposed model with other deep learning model.

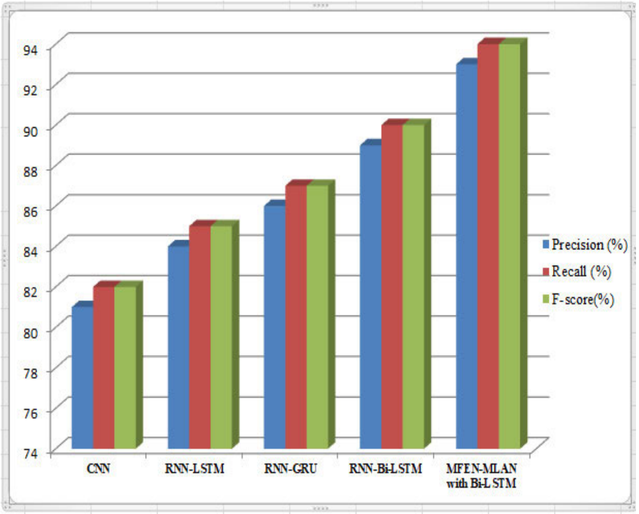

FIGURE 5. Comparison of proposed model with data augmentation.

This result illustrates that the CNN networks provide superior result when the training datasets are high.

In addition, the proposed model MFEN-MLAN with Bi-LSTM model is compared with other deep learning models on three publicly available standard datasets ICDAR2003 dataset, ICDAR2013 dataset and SVT dataset respectively and it performs well than other models, as shown in Table 4 Table 5 and Table 6.

TABLE 4. Performance analysis of MFEN-MLAN with Bi-LSTM model on ICDAR2003 dataset.

Models	Precision (%)	Recall (%)	F-score(%)
CNN	80.5	82.1	81.3
RNN-LSTM	83.5	85.2	84.3
RNN-GRU	84.2	86.3	85.2
RNN-Bi-LSTM	85.7	87.1	86.4
MFEN-MLAN with Bi-LSTM	93.5	95	94.2

TABLE 5. Performance analysis of MFEN-MLAN with Bi-LSTM model on icdar2013 dataset.

Models	Precision (%)	Recall (%)	F-score(%)
CNN	84.6	86.5	85.5
RNN-LSTM	85.2	87.8	86.5
RNN-GRU	92.6	93.6	93.1
RNN-Bi-LSTM	94.9	95.2	95.0
MFEN-MLAN with Bi-LSTM	95.2	97.8	96.5

TABLE 6. Performance analysis of MFEN-MLAN with Bi-LSTM model on SVT dataset.

Models	Precision (%)	Recall (%)	F-score(%)
CNN	80.1	82.2	81.1
RNN-LSTM	82.5	83.6	83.0
RNN-GRU	83.4	84.6	84.0
RNN-Bi-LSTM	84.9	86.8	85.8
MFEN-MLAN with Bi-LSTM	94.9	95.3	95.1

E. PERFORMANCE ANALYSIS

One of the contributions of this paper is recognizing Tamil characters from the news tickers, and there is no prior work for this problem in Tamil language. Apart from this, English character recognition has been done in natural scene images with two different datasets. To evaluate the quality of the proposed model, our proposed method is compared with ICDAR2003 (IC03) dataset, ICDAR2013 (IC13) dataset, and Street View Text (SVT) dataset, as mentioned before, along with various nature scene text recognition models [49]. The results of the experiments are compared with other related works and are depicted in Table 7. From the table, it is evident that the recommended fusion feature method exceeds the other state-of-the-art methods. Recognizing the similar structure characters such as i, l, u, and v is challenging, and it increases the recognition error rate. The F1-Score of proposed

TABLE 7. Comparison of proposed model with other models in terms of F-Score.

Model	Method	ICDAR03	ICDAR13	SVT
Anand Mishra et. al. [24]	CFR	68	-	72
Jaderberg et. al. [25]	CNN	93.1	90.8	80.7
Gomezbigorda et. al. [26]	Search Algorithm	75	83.58	54
Alsharif et. al. [27]	HMM	88	-	74
Yao et. al. [28]	Strokelet	75	-	80
Liu et. al. [42]	STAR-net	89.9	89.1	83.6
Buta et. al. [33]	FASText	-	76.8	-
Shi et. al. [19]	CRNN	91.9	89.6	82.7
Deng et. al. [34]	PixelLink	-	84.5	-
Cheng et. al. [43]	arbitrarily-oriented text	91.5	-	82.8
Minghui et. al. [44]	2D perspective	-	91.5	86.4
Jaderberg et. al. [32]	Deep structure learning	89.6	81.8	71.7
Lyu et. al. [35]	Corner	-	85.8	-
Heng et. al. [45]	Rectification module	-	96.0	90.0
Liao et. al. [36]	Text boxes++	-	80.0	-
Gao et. al. [46]	Fully convolutional sequence modeling	89.2	88.0	82.7
Cao et. al. [37]	FDTA	-	88.7	-
Mokayed et. al. [38]	Fuzzy	-	-	90
Shi et. al. [47]	Flexible rectification	94	91.8	93
Van et. al. [39]	MPT	-	-	88.4
Liu et. al. [40]	Attention	-	89	87
Gandhewar et. al. [41]	RCNN	80.1	84.1	71
Proposed Model	MFEN-MLAN with Bi-LSTM	94.2	96.5	95.1

TABLE 8. Performance comparison of networks in terms of F-score.

Models	TNT	ICDAR03	ICDAR13	SVT
Baseline	84.1	83.6	82.9	85.1
Baseline + MFEN	86.4	85.2	84.6	88.9
Baseline+ MLAN	88.2	89.4	87.8	91.4
Baseline + MFEN + MLAN	94	94.2	93.1	95.1

model on ICDAR03 dataset is 94.2%, ICDAR13 dataset is 96.5% and SVT dataset is 95.1% which is reasonably higher than the other works.

The model consists of two different networks 1. MFEN 2. MLAN. For evaluating the efficiency of these networks, the model is evaluated with four different variants. The variant uses the baseline network alone; the second variant uses the baseline network along with MFEN; third variant uses the baseline network with MLAN network and final variant uses the baseline network with MFEN and MLAN networks. Out of all the variants, the variant that used both the networks (MFEN and MLAN) yields the best results. All the variants use the same network tuning parameters. Table 8 represents the comparison of all the variants in terms of F1-Score and it is shown in Figure 6. From the table, it is clear that the proposed feature fusion methodology yields a better F1-Score. The experiments were carried out with 4GB of random access memory and on an Intel Core i3 2.00 GHz CPU. The mean time for training and testing for the given datasets is shown in Table 9.

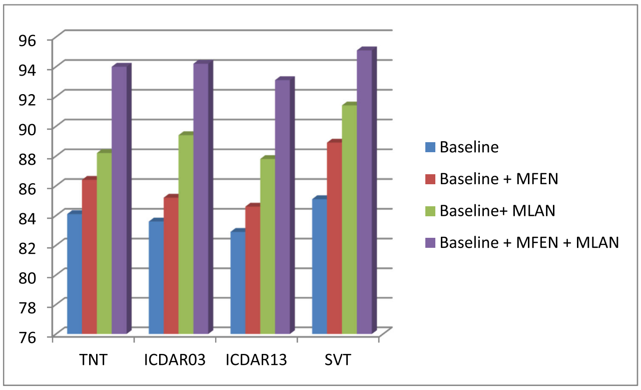

FIGURE 6. Performance comparison of networks in terms of F-score.

TABLE 9. Average running time in seconds for the proposed model.

Dataset	On training	On testing
ICDAR 03	0.035	0.12
ICDAR 13	0.036	0.12
SVT	0.037	0.11
Tamil News ticker (TNT)	0.036	0.13

V. CONCLUSION

The proposed research work presents an improved Convolutional Neural Network architecture that integrates Manifold Feature Extraction Neural network(MFEN) and Multi-layer Feature Aggregation Neural network (MLAN). For combining the features of the different convolutional layers in the neural network, MFEN network is used for extracting the significant features from the input images; these features with high-level layer are aggregated in MLAN. Thus Convolutional Neural Network (CNN) effectively identifies features which are fed into the Bidirectional Long Short Term Memory (Bi-LSTM) model for recognizing the text label sequences from the input images and the decoder layer is used to translate the frame sequence into appropriate final result for the text prediction on image. To evaluate the performance of our proposed model, Tamil news tickers images dataset is created, which is considered as one of the contributions of this research work and this is the first benchmark dataset for Tamil news tickers images. The accuracy of detecting Tamil text on images is compared with other well-known models, and the data augmentation method was applied to enlarge the sample images of training set in our Tamil news tickers image dataset. In addition, our proposed model is also evaluated on other English character dataset such as SVT dataset, ICDAR03 dataset, ICDAR13 dataset and their results are discussed in this research work. Finally, for recognizing the text in natural scene images, our proposed model produced promising results when compare to the other advanced methods.

REFERENCES

[1] A. A. Chandio, M. Asikuzzaman, and M. R. Pickering, ''Cursive character recognition in natural scene images using a multilevel convolutional neural network fusion,'' *IEEE Access*, vol. 8, pp. 109054–109070, 2020, doi: 10.1109/ACCESS.2020.3001605.

[2] S. Y. Arafat and M. J. Iqbal, ''Urdu-text detection and recognition in natural scene images using deep learning,'' *IEEE Access*, vol. 8, pp. 96787–96803, 2020, doi: 10.1109/ACCESS.2020.2994214.

[3] Y. S. Chernyshova, A. V. Sheshkus, and V. V. Arlazarov, ''Two-step CNN framework for text line recognition in camera-captured images,'' *IEEE Access*, vol. 8, pp. 32587–32600, 2020, doi: 10.1109/ACCESS.2020.2974051.

[4] J. Xu, W. Ding, and H. Zhao, ''Based on improved edge detection algorithm for English text extraction and restoration from color images,'' *IEEE Sensors J.*, vol. 20, no. 20, pp. 11951–11958, Oct. 2020, doi: 10.1109/JSEN.2020.2964939.

[5] S. Anbukkarasi and S. Varadhaganapathy, ''A novel approach for handwritten Tamil character recognition system,'' *J. Adv. Res. Dyn. Control Syst.*, vol. 12, pp. 1489–1495, Mar. 2020.

[6] M. A. Panhwar, K. A. Memon, A. Abro, D. Zhongliang, S. A. Khuhro, and S. Memon, ''Signboard detection and text recognition using artificial neural networks,'' in *Proc. IEEE 9th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC)*, Beijing, China, Jul. 2019, pp. 16–19, doi: 10.1109/ICEIEC.2019.8784625.

[7] S. Attigeri, ''Neural network based handwritten character recognition system,'' *Int. J. Eng. Comput. Sci.*, vol. 7, no. 3, pp. 23761–23767, Mar. 2018.

[8] T. Intasuwan, J. Kaewthong, and S. Vittayakorn, ''Text and object detection on billboards,'' in *Proc. 10th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Bali, Indonesia, Jul. 2018, pp. 6–11, doi: 10.1109/ICITEED.2018.8534879.

[9] B. U. Tayyab, M. F. Naeem, A. Ul-Hasan, and F. Shafait, ''A multi-faceted OCR framework for artificial Urdu news ticker text recognition,'' in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Vienna, Austria, Apr. 2018, pp. 211–216, doi: 10.1109/DAS.2018.83.

[10] A. A. Prakash and S. Preethi, ''Isolated offline Tamil handwritten character recognition using deep convolutional neural network,'' in *Proc. Int. Conf. Intell. Comput. Commun. Smart World (ICSW)*, Erode, India, Dec. 2018, pp. 278–281, doi: 10.1109/I2C2SW45816.2018.8997144.

[11] N. R. Prashanth, B. Siddarth, A. Ganesh, and V. N. Kumar, ''Handwritten recognition of Tamil vowels using deep learning,'' *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 263, Nov. 2017, Art. no. 052035, doi: 10.1088/1757-899X/263/5/052035.

[12] X. Ren, Y. Zhou, Z. Huang, J. Sun, X. Yang, and K. Chen, ''A novel text structure feature extractor for Chinese scene text detection and recognition,'' *IEEE Access*, vol. 5, pp. 3193–3204, 2017, doi: 10.1109/ACCESS.2017.2676158.

[13] O. Zayene, J. Hennebert, S. M. Touj, R. Ingold, and N. E. B. Amara, ''A dataset for Arabic text detection, tracking and recognition in news videos—AcTiV,'' in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Tunis, Tunisia, Aug. 2015, pp. 996–1000, doi: 10.1109/ICDAR.2015.7333911.

[14] M. D. A. Asif, U. U. Tariq, M. N. Baig, and W. Ahmad, ''A novel hybrid method for text detection and extraction from news videos,'' *Middle-East J. Sci. Res.*, vol. 19, pp. 716–722, Jan. 2014.

[15] A. Ali, M. Pickering, and K. Shafi, ''Urdu natural scene character recognition using convolutional neural networks,'' in *Proc. IEEE 2nd Int. Workshop Arabic Derived Script Anal. Recognit. (ASAR)*, London, U.K., Mar. 2018, pp. 29–34, doi: 10.1109/ASAR.2018.8480202.

[16] Q. Liang, S. Xiang, Y. Wang, W. Sun, and D. Zhang, ''RNTR-Net: A robust natural text recognition network,'' *IEEE Access*, vol. 8, pp. 7719–7730, 2020, doi: 10.1109/ACCESS.2020.2964148.

[17] A. Graves, S. Fernandez, F. Gomez, and J. Schmidhuber, ''Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks,'' in *Proc. 23rd Int. Conf. Mach. Learn.*, Jun. 2006, pp. 369–376.

[18] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[19] B. Shi, X. Bai, and C. Yao, ''An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017, doi: 10.1109/TPAMI.2016.2646371.

[20] K. Wang, B. Babenko, and S. Belongie, ''End-to-end scene text recognition,'' in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1457–1464, doi: 10.1109/ICCV.2011.6126402.

[21] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, ''The marginal value of adaptive gradient methods in machine learning,'' in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4151–4161.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ''Learning representations by back-propagating errors,'' *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[23] [Online]. Available: https://keras.io/

[24] A. Mishra, K. Alahari, and C. Jawahar, ''Scene text recognition using higher order language priors,'' in *Proc. Brit. Mach. Vis. Conf.*, Surrey, U.K., 2012, p. 127, doi: 10.5244/c.26.127.

[25] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, ''Reading text in the wild with convolutional neural networks,'' *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 1–20, May 2015, doi: 10.1007/s11263-015-0823-z.

[26] L. Gómez and D. Karatzas, ''TextProposals: A text-specific selective search algorithm for word spotting in the wild,'' *Pattern Recognit.*, vol. 70, pp. 60–74, Oct. 2017, doi: 10.1016/j.patcog.2017.04.027.

[27] O. Alsharif and J. Pineau, ''End-to-end text recognition with hybrid HMM maxout models,'' 2013, *arXiv:1310.1811*.

[28] C. Yao, X. Bai, B. Shi, and W. Liu, ''StrokeLets: A learned multi-scale representation for scene text recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 4042–4049, doi: 10.1109/CVPR.2014.515.

[29] F. Gil and S. Osowski, ''Fusion of feature selection methods in gene recognition,'' *Bull. Polish Acad. Sci., Tech. Sci.*, Jan. 2021, Art. no. e136748, doi: 10.24425/bpasts.2021.136748.

[30] H. Meng, T. Yan, H. Wei, and X. Ji, ''Speech emotion recognition using wavelet packet reconstruction with attention-based deep recurrent neutral networks,'' *Bull. Polish Acad. Sci., Tech. Sci.*, vol. 69, no. 1, pp. 1–12, 2021, doi: 10.24425/bpasts.2020.136300.

[31] A. Osowska-Kurczab, T. Markiewicz, M. Dziekiewicz, and M. Lorent, ''Multi-feature ensemble system in the renal tumour classification task,'' *Bull. Polish Acad. Sci., Tech. Sci.*, vol. 69, no. 3, 2021, Art. no. e136749, doi: 10.24425/bpasts.2021.136749.

[32] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, ''Deep structured output learning for unconstrained text recognition,'' 2014, *arXiv:1412.5903*.

[33] M. Buta, L. Neumann, and J. Matas, ''FASText: Efficient unconstrained scene text detector,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1206–1214, doi: 10.1109/ICCV.2015.143.

[34] D. Deng, H. Liu, X. Li, and D. Cai, ''PixelLink: Detecting scene text via instance segmentation,'' in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 6773–6780, doi: 10.1609/aaai.v32i1.12269.

[35] P. Lyu, C. Yao, W. Wu, S. Yan, and X. Bai, ''Multi-oriented scene text detection via corner localization and region segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7553–7563.

[36] M. Liao, B. Shi, and X. Bai, ''TextBoxes++: A single-shot oriented scene text detector,'' *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3676–3690, Aug. 2018, doi: 10.1109/TIP.2018.2825107.

[37] Y. Cao, S. Ma, and H. Pan, "FDTA: Fully convolutional scene text detection with text attention," *IEEE Access*, vol. 8, pp. 155441–155449, 2020, doi: 10.1109/ACCESS.2020.3018784.

[38] H. Mokayed, P. Shivakumara, R. Saini, M. Liwicki, L. C. Hin, and U. Pal, "Anomaly detection in natural scene images based on enhanced fine-grained saliency and fuzzy logic," *IEEE Access*, vol. 9, pp. 129102–129109, 2021, doi: 10.1109/ACCESS.2021.3103279.

[39] D. N. Van, S. Lu, X. Bai, N. Ouarti, and M. Mokhtari, "Max-pooling based scene text proposal for scene text detection," in *Proc. 14th Int. Conf. Document Anal. Recognit. (ICDAR)*, Kyoto, Japan, 2017, pp. 1295–1300, doi: 10.1109/ICDAR.2017.213.

[40] C. Liu, Y. Zou, and W. Guan, "Hierarchical feature fusion with text attention for multi-scale text detection," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Shanghai, China, Nov. 2018, pp. 1–5.

[41] N. Gandhewar, S. R. Tandan, and R. Miri, "Deep learning based framework for text detection," in *Proc. 3rd Int. Conf. Intell. Commun. Technol. Virtual Mobile Netw. (ICICV)*, Feb. 2021, pp. 1231–1236, doi: 10.1109/ICICV50876.2021.9388529.

[42] W. Liu, C. Chen, K.-Y. Wong, Z. Su, and J. Han, "STAR-net: A SpaTial attention residue network for scene text recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2016, pp. 1–18.

[43] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, and S. Zhou, "AON: Towards arbitrarily-oriented text recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5571–5579.

[44] M. Liao, J. Zhang, Z. Wan, F. Xie, J. Liang, P. Lyu, C. Yao, and X. Bai, "Scene text recognition from two-dimensional perspective," in *Proc. AAAI Conf. Artif. Intell.*, Jan. 2019, pp. 8714–8721.

[45] H. Heng, P. Li, T. Guan, and T. Yang, "Scene text recognition via context modeling for low-quality image in logistics industry," *Complex Intell. Syst.*, pp. 1–20, Nov. 2022, doi: 10.1007/s40747-022-00916-1.

[46] Y. Gao, Y. Chen, J. Wang, M. Tang, and H. Lu, "Reading scene text with fully convolutional sequence modeling," *Neurocomputing*, vol. 339, pp. 161–170, Apr. 2019.

[47] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: An attentional scene text recognizer with flexible rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2035–2048, Sep. 2019, doi: 10.1109/TPAMI.2018.2848939.

[48] W. Xue, Q. Li, and Q. Xue, "Text detection and recognition for images of medical laboratory reports with a deep learning approach," *IEEE Access*, vol. 8, pp. 407–416, 2020, doi: 10.1109/ACCESS.2019.2961964.

[49] P. Keserwani, R. Saini, M. Liwicki, and P. P. Roy, "Robust scene text detection for partially annotated training data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8635–8645, Dec. 2022, doi: 10.1109/TCSVT.2022.3194835.

[50] P. Dai, Y. Li, H. Zhang, J. Li, and X. Cao, "Accurate scene text detection via scale-aware data augmentation and shape similarity constraint," *IEEE Trans. Multimedia*, vol. 24, pp. 1883–1895, 2022, doi: 10.1109/TMM.2021.3073575.

[51] T. Ali, M. F. H. Siddiqui, S. Shahab, and P. P. Roy, "GMIF: A gated multiscale input feature fusion scheme for scene text detection," *IEEE Access*, vol. 10, pp. 93992–94006, 2022, doi: 10.1109/ACCESS.2022.3203691.

[52] T. Guan, "Industrial scene text detection with refined feature-attentive network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 6073–6085, Sep. 2022, doi: 10.1109/TCSVT.2022.3156390.

[53] M. Liao, Z. Zou, Z. Wan, C. Yao, and X. Bai, "Real-time scene text detection with differentiable binarization and adaptive scale fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 919–931, Jan. 2023, doi: 10.1109/TPAMI.2022.3155612.

S. ANBUKKARASI received the B.Tech. degree from the Kongu Engineering College, Erode, the M.E. degree from the Vidyaa Vikas College of Engineering and Technology, and the Ph.D. degree in information and communication engineering from Anna University, Chennai, in 2022. She is currently an Assistant Professor with the Department of Computer Science and Engineering, Kongu Engineering College. She has published many research papers in various national and international journals and conferences. Her research interests include natural language processing, deep learning, machine learning, and sentiment analysis. She has very depth knowledge of her research areas.

VEERAPPAMPALAYAM EASWARAMOORTHY SATHISHKUMAR received the bachelor's degree in information technology from the Madras Institute of Technology, Anna University, in 2013, the master's degree in biometrics and cyber security from the PSG College of Technology, in 2015, and the Ph.D. degree from Sunchon National University, in 2021. He was a Research Associate with VIT University, from 2015 to 2017. He was a Postdoctoral Researcher with the Department of Industrial Engineering, Hanyang University, Seoul, South Korea. In 2021, he was an Assistant Professor with the Department of Computer Science and Engineering, Kongu Engineering College. He is currently a Postdoctoral Researcher with the Department of Software Engineering, Jeonbuk National University, South Korea. He has published more than 60 research papers in reputed journals and conferences. His research interests include data mining, big data analytics, cryptography, digital forensics, artificial intelligence, and computational chemistry. He received the South Korea's prestigious Global Korean Scholarship for pursuing the Ph.D. degree. He is a reviewer of more than 200 journals. He has reviewed more than 2000 research articles. He is currently serving as an Academic Editor for the journals *PLOS ONE* and *Journal of Healthcare Engineering*.

C. R. DHIVYAA received the B.Tech. degree in information technology from the Velalar College of Engineering and Technology, Anna University, Chennai, in 2010, the M.Tech. degree in information technology from the Sasurie College of Engineering, Anna University, in 2012, and the Ph.D. degree in information and communication engineering from Anna University, in 2019. She is currently an Assistant Professor with the Department of Computer Science and Engineering, Kongu Engineering College, Erode. She has published many research papers in various national and international journals and conferences. Her current research interests include computer vision, image processing, medical ima'ging, machine learning, and pattern recognition. She has very depth knowledge of her research areas.

JAEHYUK CHO received the Ph.D. degree in computer science from Chung-Ang University, South Korea, in 2011, with a focus on mobile and embedded computing systems. He was a Professor with the Department of Electronic Engineering, Soongsil University. He was a National Research and Development Program Project Manager with the Korea Institute of Science and Technology Evaluation and Planning (KISTEP), Seoul. He was a Senior Researcher with LG CNS, Seoul. He is currently a full-time Professor with the Department of Software Engineering, Jeonbuk National University, Jeonju, South Korea. His research interests include applied AI, data process, big data of sensors, the IoT, smart city, and SW platform systems.