International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 08 | Aug 2022 www.irjet.net p-ISSN: 2395-0072

Deep Learning in Text Recognition and Text Detection : A Review

Brindha Muthusamy[1], Kousalya K[2]

[1]Dept. of Computer Science and Engineering, Kongu Engineering College, Erode
[3]Professor, Dept. of Computer Science and Engineering, Kongu Engineering College, Erode

---

Abstract - *Detecting text in natural situations is a difficult task that is more difficult than extracting the text from those natural images in which the background and foreground are clearly separated and every character is isolated from the images. Text in the landscapes which is nature may occur in a range of states like a text in dark with the background light and vice versa, with a broad diversity of fonts, even for letters of the same word, sections of words can be overlapped by environment objects, making detection of these parts impossible. Deep learning which is a subset of Machine learning employs a neural network, a technique that replicates how the brain analyses data. An Optical Character Recognition engine has two parts: i)Text recognition and ii)Text detection. The process of locating the sections of text in a document is known as text detection. Since the different documents (invoices, newspapers, etc.) have varied structures, this work has historically proved difficult. A text recognition system, on the other hand, takes a portion of a document containing text (a word or a line of text) and outputs the associated text. Both text detection and text recognition have shown considerable promise with deep learning algorithms.*

Keywords: Deep learning, Convolutional Neural Network, text detection, text classification, Optical Character Recognition, optimization

I. INTRODUCTION

Deep learning algorithms learn about the image by passing through each neural network layer. For applications like machine translation and image searches, OCR mechanism is being used to recognise text within photographs. The method for recognising the text/characters from a photo, sends the features extracted text from an image/photo to the classifier which is trained to distinguish individual characters that is similar to object recognition. Recognition of text, on the other hand, looks at the text as a group of meaningful characters rather than single characters. A text string can be recognized by clustering the similar characters, which means that every character in the text must separate for recognition **[21]**. Alternatively, in the instance of classification, train the network on a labelled datasets in order to categorize the samples in the datasets.

Convolutional Neural Network (CNN) was applied to extract the features for character recognition performance. A brief review on text detection and text recognition are shown in Table 2. The CNN model is trained in three steps : (i) Train the CNN model from the scratch, (ii) Using a transfer learning method for exploiting the features from a pretrained model on larger datasets and, (iii) Transfer learning and fine-tuning the weights of an CNN architecture. CNN architecture consists mostly of four layers. (i) Convolutional layer [Conv]; (ii) Pooling layer; (iii) Fully connected layer; and (iv) Rectified Linear units **[4]**.

Artificial Neural Network (ANN) is being trained to extract information through the deep-learning-based technologies from an image. VGG, ResNet, MobileNet, GoogleNet, Xception, and DenseNet are examples of CNN architecture where several convolution layers are utilized to extract the features from the images **[21]**. Handwritten texts are difficult while reading text from photographs or recognising that has attained a lot of attention. Most systems have two important components, (i) Text detection; and (ii) Text recognition. Text detection is a technique where the text instances from the images can be predicted and localized. The text recognition is done by autoencoder which is the process of decoding the text into the machine-readable format.

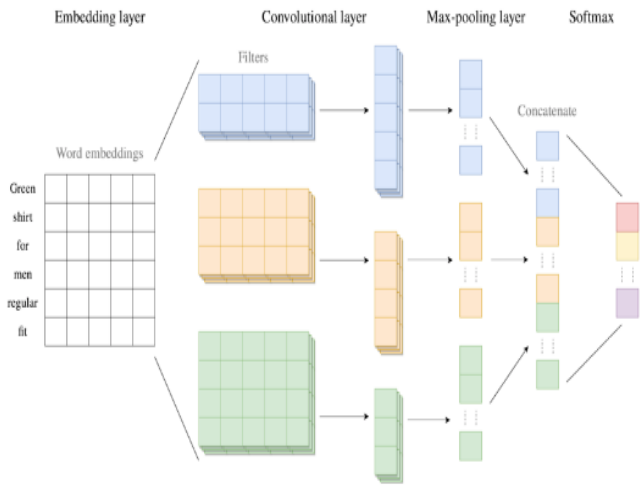

Figure 3: General CNN architecture for text classification

Figure 1 CNN architecture for text classification

II. TEXT DETECTION AND RECOGNITION USING DEEP LEARNING

2.1 Text detection

The local features can be extracted from CNN through the training of character pictures and sub-regions based on the characteristics of an individual's handwriting. To train CNN, randomly selected instances of an image in the training sets are used, and the local features extracted from the image from these instances are aggregated in order to generate the global features. The process of randomly sampling instances is repeated for every training epoch. This results in the increase of training the patterns for training the CNN for text independent writer identification **[1].**

Deep learning can constantly learn by examining data and finding patterns and classifying images. For picture categorization, language transaction and character recognition, deep learning is applied. Deep Neural Networks are networks with multiple layers that can perform complex operations on images, sound, and text, such as representation and abstraction. It can be used for any type of recognition problem. The basic goal is only for computers to learn without human intervention and modify their activities accordingly.

2.2 Text recognition

The recognition can be done by many techniques. It involves Convolutional Neural Network(CNN), Semi Incremental Method, Incremental Method, Line and Word Segmentation method etc. One of the most effective and prominent ways of handwriting recognition is Convolutional Neural Network (CNN). The most prevalent application of CNN is in image analysis. Artificial neurons are used in CNN **[11]**. CNN can recognise images and videos, classify images, analyse medical images, perform computer vision, and process natural language.

CNN or ConvNets architecture has made significant contributions to the analysis of images. CNN is defined as , 1)A convolution tool that separates and identifies the distinct characteristics of the image for analysis in a process called Feature Extraction, which is part of the CNN architecture. 2) A fully connected layer makes use of the output of the convolution process in order to forecast the image's class using the information acquired in earlier stages.

CNN extracts the features from handwritten text images of numerous characters by the end-to-end method based on deep learning and combines these extracted features to recognize the writer's specific data during lowering of character's class-specific features. To create n-tuple images, a single writer's handwritten squared images are sampled randomly. Second, every image from n-tuple is sent into a local feature extractor (CNN). CNN can extract text-independent writer specific properties by employing new techniques to structure the training samples as n tuple pictures. Finally, a global feature aggregator aggregates the retrieved local features in various ways, such as using the average or maximum. Finally, the combined characteristics are sent into a softmax classifier (fully connected with an Nfc unit which is equal to the number of writers) for predicting **[1]**.

III. PRELIMINARY STEPS FOR TEXT DETECTION AND RECOGNITION

3.1 Datasets

The datasets are collected from various ImageNet dataset like, MNIST, NIST, IAM dataset, IFN/ENIT, and so on. The different dataset on different languages like, English, Arabic, Bangla handwritten character datasets are taken into consideration. The detailed view of these datasets are shown in table 2.

3.2 Image Pre-processing

Pre-processing is a technology which transforms the raw data to useful and efficient data/information. This process consists of different operations which are used to perform on input images and in this process, the images are reshaped. The rearrangement of the form of the data is done without changing the contents of the data. Different kinds of arrangements are done in this process according to the parameters which are needed to carry on till further process **[1]**. Image pre-processing is nothing but to remove irrelevant data and delete the duplicate data from the databases. The text in the image can be in different styles, fonts and so on. In data pre-processing, the data are processed in different ways, they are:

- **Data cleaning** : Remove noise and data inconsistency from raw data. That is, the dataset which does not belong to the dataset comes, this data cleaning process removes these datasets. It is done by filtering the dataset and handling the missing data. It ensures the quality of data. In other words, it only removes the unwanted data but maintains the originality of the data.

- **Data integration** : Data integration collects data from various multiple resources and combines it to form coherent data and also supports the consolidated perspective of the information. That is, it merges data from various sources into a coherent datastore (data warehouse). These data are stored and maintained for future use. It may be in the form of documents where the relevant data are stored in the different documents.

- **Data reduction** : Data reduction is the process of reducing the data size by instances, aggregation, eliminating irrelevant data/features or by clustering. Data reduction can increase the storage capacity and the cost is reduced. It does not lose any data, instead it maintains the originality of the data. Data reduction process is done with the help of data compression. Data compression is a technique to compress the data in a reduced form of data which can reduce the storage space.

- **Data transformation** : Data transformation is used to convert one form of format to another format. It is also known as data munging or data wrangling. Otherwise called Normalization. The data points in the scatter plot should be linear. If the points are in the form of a curve, it is difficult to calculate the accuracy which affects the performance of the model. This curve can be converted into a linear line by scaling the model in the range of 0 and 1. Activation functions are used when the model shows non-linearity in their respective model.

3.3 Normalization

Normalization is the process of organizing data in the database. This process involves table creation and establishing relationships between these tables. It protects the data as well as makes the database more flexible by eliminating redundancy and inconsistent dependency. Redundant data takes more space and is complicated to maintain. The inconsistent dependent data can make it difficult to access the data since the path to find the data is either missing or broken. Hence, Normalization is more important since it can reduce these irrelevant data and data inconsistency and it can handle the missing data as well to make the database more flexible. Normalization includes three stages of normalization steps where each stage generates the table. Each table stores the relevant data which does not include duplicate data or miss any data. The three steps involved in the normalization process are as follows,

3.3.1 First Normal form

The First Normal Form (1NF) sets the fundamental rules for database normalization which relates to the single table in the relational database model. The steps involved in the First Normal Form are :

- Every column in the table are unique
- Separate tables are created for every relevant set of data
- Each table must be identified with a unique column or the concatenated columns are called with the primary key

- Neither row or column of the table are duplicated
- No row or column that intersects in the table contain a NULL value
- No row or column that intersects in the table can have multi-valued fields

3.3.2 Second Normal Form

The Second Normal Form (2NF) follows the First Normal Form . The basic requirements of Second Normal Form for organizing the data include:

- No redundancy of data. All the data is stored in only one place.
- Data dependencies in Second Normal Form are logical. That is, all the related data items are stored together which is useful for easy access.

3.3.3 Third Normal Form

The Third Normal Form (3NF) is the combination of both First Normal Form and Second Normal Form (1NF+2NF). The main benefits of this Third normal form are:

- Reduces duplication of the data and achieves data integrity in a database.
- Useful to design a normal relational database
- 3NF are independent of anomalies of deletion, updation, and insertion
- It ensures losslessness and prevention of the functional dependencies

3.4 Feature extraction

Text extraction or Feature extraction from an image is a method of extracting text from a photograph using machine learning techniques. Text extraction is also known as text localization. It is used for text detection and localization which helps for text recognition. Text localization is the process which is used to develop a computer system (AI) to automatically recognize and read the text from the images. In deep learning models, the feature extraction process is done automatically since the models are pre-trained. It is done by text classification.

3.4.1 Text Classification

Text classification is also known as text tagging or text categorization. Texts are always unstructured in handwritten words or characters. Hence, these texts are categorized into an organized group. This process is difficult since it takes more time and is a little expensive. By using Natural Language Processing (NLP) which can be used as a text classifier, the texts can be categorized and can be converted into structural textual data which is easy to understand, cost effective and is more scalable. Natural Language Processing can automatically analyse and understand the type of character and it will assign a set of pre-defined tags (Pre-trained image datasets) and can be categorized based on its characteristics.

3.5 Text detection

In the testing stage, text detection is accomplished by segmenting photos. A series of text/sub-images of individual text is divided from a full image. Edge detection and the space between the different characters are used to segment the image. Following segmentation, the sub-divided portions are labelled and processed one at a time. This labelling is used to determine the total number of characters in an image. After that, each sub picture is scaled (70x50) and normalized in relation to itself. This aids in the extraction of image quality attributes **[9]**. Text detection is a method in which the model is given an image and the text region is detected by building a bounding box around it. Text recognition is carried out by further processing the discovered textual sections in order to recognise the text.

3.5.1 Training the dataset

Deep learning model is built while feeding data to a deep neural network (DNN) to "train" in order to do a specific AI task (such as image classification or speech to text conversion). The hidden layer in the neural network is used for backpropagation which helps to improve the performance of the training model. Hence, during the training process, known data is fed into the DNN, and this DNN generates a prediction on what the data represents.

Image Classification (used to classify the type of an object in an image).

- Input: A photograph, for example, is a single-object image.
- Output: A designation for a class (one or more integers that are mapped to class labels).

3.5.2 Validating the dataset

Validation data introduces new data into the model that it hasn't assessed before during training. Validation data serves as the first test against unknown data, allowing data scientists to assess how well the model predicts new data. Although not all data scientists use validation data, it can be useful in optimizing hyperparameters, which influence how the model evaluates data.

Object Localization (locate the existence of objects in an image and use a bounding box to represent their location).

- Input: A photograph, for instance, is an image featuring one or more objects.
- Output: A set of bounding boxes (or more) .

3.5.3 Testing the dataset

Testing data once the model has been developed confirms that it can make accurate predictions. The testing data should be unlabelled if the training and validation data include labels to track the model's performance metrics. Test data is the verification of an unknown dataset to make sure that the machine learning algorithm was trained properly.

Object Detection (detect the presence of objects/characters in an image with a bounding box, as well as the types or classes of those objects).

- Input: A photograph, for instance, is an image featuring one or more objects.
- Output: A class label for each bounding box, as well as one or more bounding boxes (specified by a point, width, and height).

3.6 Text recognition

Text recognition, commonly known as OCR (Optical Character Recognition) is a type of computer vision problem which involves converting images/photos of digital or hand-written character into a machine-readable text that the computer can process, save, and edit as a text file or as part of data entry and manipulation software. It can recognize the text from any documents, handwritten characters, and so on by pre-training the images using deep learning concepts and its architectures. In other words, text recognition is the process where the image text is converted into a recognisable and readable text. The recognition of the detected text is done by various deep neural networks.

Year	CNN architecture	No. of layers	No. of parameter
2014	VGG-16	16 layers	138 million
2015	ResNet	34 layers	23 million
2014	GoogleNet	22 layers	7 million
2016	Xception	71 layers	~26 million
	DenseNet	121 layers	20 million

Table 1 CNN architecture

Some of the advantages of Optical Character Recognition

- Quick data retrieval

- Easy to extract relevant data
- Low cost for usage like copying, editing and so on
- High accuracy to detect and recognize
- Increase the storage capacity

IV. DEEP NEURAL NETWORKS FOR TEXT DETECTION AND RECOGNITION

Deep learning is a domain of study that aims to copy the functioning of human brains in order to process data and make decisions. Deep learning is also known as Deep Neural Network (DNN) or Deep Neural Learning. The deep learning model for Optical Character Recognition uses two types of neural network architecture. They are i)Convolutional Neural Network (CNN); ii)Artificial Neural Network (ANN).

4.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a neural network which has three layers. They are: (i) Input layer, (ii)Hidden layer ;and (iii) Output layer.

Optical Character Recognition (OCR) considers the optical image of a character as an input and provides the corresponding recognisable character as an output. The trained CNN is used to extract the features of the image. CNN classifiers along with other classifiers can be combined and give the best result for classification. That is, the accuracy and efficiency of the classification can be improved. The hidden layer uses the input to perform feed-forward and backpropagation in order to improve the accuracy and reduce the error rate. Convolutional Neural Network and Error Correcting Output Code (CNN +ECOC) where CNN for feature extraction and ECOC for classification are combined to obtain Optical Character Recognition. This method gives high accuracy for handwritten characters. It is trained and validated with NIST dataset **[46].**

4.2 Artificial Neural Network (ANN)

Artificial Neural Network can be used to recognise the image with a single character and further include many characters for classification. The classification process can be done in two phases: i) Feature preprocessing - read each character and convert it into binary image and scan by all four sides (left, tp, right, bottom) and ii) a) Training the neural network and b) Testing the neural network with the datasets. Here, the training sets are used to learn how to remove noise from the data **[50]**.

4.3 Recurrent Neural Network (RNN)

Recurrent Neural Network is based on the sequential form of data. Recurrent neural network is most suitable for the text classification which can recognize the whole sentence or sequential set of words. When the words or characters are in sequence, this recurrent neural network can predict the next word of the sequence. Hence handwritten characters can be predicted more accurately while using recurrent neural networks. To achieve the prediction accuracy, RNN does not require a dataset in the form of labelled data (need not be supervised learning). Recurrent neural networks are capable of working the temporal information. The disadvantage of RNN is that it will raise the vanishing gradient problem. The training of datasets is more complex and hence more difficult. It is also difficult to process the long sequential characters.

LEARNING RATE IN DEEP LEARNING ALGORITHM

The learning rate refers to the number of times the images are trained. In deep learning neural networks, the Stochastic Gradient Descent is used. The learning rate refers to the parameters or hyperparameter which controls how many times the dataset is trained in order to reduce error rate and improve accuracy which improves the performance of the model. While training this dataset, backpropagation technique is used where the weights are updated each time in order to get better performance of the neural network model. This learning rate affects the performance of the model when it reaches the local minima. Hence, it is important to adjust the learning rate from high to low to slow down once the model reaches the optimal solution during training.

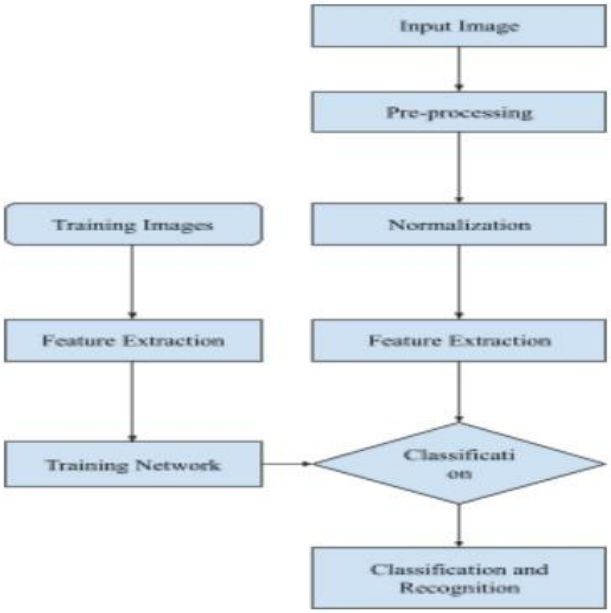

Figure 2 Process of CNN model

V. CHALLENGES IN TEXT DETECTION AND RECOGNITION

1. A wide range of natural photos - Characters have different font-styles, sizes, and colours, as well as distinctive font alignment in the natural images. The language also varies depending on the state or country or even other region.

2. Background complexity - Natural photographs' backgrounds may not be as distinct. The background comprises grasses, bricks, pebbles, and sign boards, making text identification more difficult.

3. Factors influencing inference - Blurring, noise, and low resolution of input images are the key inference variables. The photos may have blurred text that cannot be recognized and read the text accurately.

4. Poor handwriting - Handwritten characters may not be clear due to various styles which may not be readable (since handwriting varies per individual).

5. Size of characters - A person's handwriting varies for every individual which is why some of the characters are not easily recognisable.

6. Language identification - The text in traffic signs and advertising panels may be in different languages that are not known by foreigners. The artificially added text on an image like watermarks or subtitles are also difficult to recognize especially when they are in foreign language.

7. The text in the videos are very difficult to read since the text is moving continuously at a particular speed. Some of these texts may or may not be clear which is very difficult to recognize accurately.

VI. DEEP LEARNING ALGORITHMS

6.1 Convolutional Neural Network (CNN)

To perform convolutional operations, CNNs process input by passing it through many layers and extracting features. The Rectified Linear Unit (ReLU) that outlasts to fix the feature map makes up the Convolutional Layer. These feature maps are rectified into the next feed using the pooling layer. Pooling is a down-sampled method which reduces the dimension of the feature map. The resulting 2-D arrays are made up of single, long, continuous, and linear vectors that have been flattened in the map. Fully-connected layer is the next layer, which takes a flattened matrix or 2-D array obtained from the pooling layer as an input and then classifies the image.

6.2 Long-Short Term Memory Networks(LSTMs)

LSTMs are proven which outperforms the conventional recurrent neural networks and it suffers from the fading gradient problem, when modelling long-term dependencies. The CTC layer, which comes after the LSTM layers [figure 3], gives the feature sequences with the ground truth transcription during the training and decoding the LSTM layer's outputs during evaluation to create the predicted transcription. From the start to finish, the system is taught through feeding the text lines, pictures and ground truth transcription (in UTF-8) **[5].**

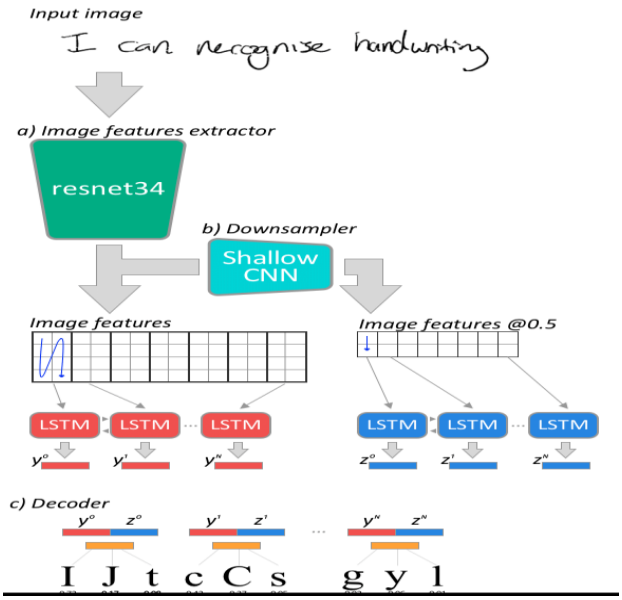

Figure 3 CNN-BiLSTM architecture

6.3 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks is mainly used for feed-forward techniques. RNN can work both parallelly and sequentially. During computation. A number of sequence learning challenges have been successfully solved using Recurrent Neural Networks (RNN) and RNN variations (Bi-directional LSTMs and MDLSTM). According to several studies, LSTM outperforms HMMs on such tasks.

6.4 Generative Adversarial Networks (GANs)

GANs are the form of neural network architecture which enables the deep learning models to learn and capture the training data distribution, which provides the generation of new data instances based on these distributions. GANs are used for unsupervised machine learning to train the two models parallely. It gives the training data which is similar to the original data. A discriminator and a generator are often included in GANs **[51].**

6.5 Multi-Layer Perceptron (MLPs)

The MLPs is a supervised learning convolutional artificial neural network that reduces error by continuously computing and updating all the weights in the network. In the first phase, which is a feed-forwarding phase, the trained data is delivered to the output layer, after that the output and the desired targets (errors) are back-propagated to update the network's weight in the next phase. The Adam optimiser was employed to improve the performance of MultiLayer Perceptron **[22].**

Adam optimizer : Adam optimizer is an extension of stochastic gradient descent that uses deep learning applications for computer vision and Natural Language Processing (NLP). Adam optimizer is used to remove vanishing gradient problems. The computation time is faster and it requires only a few parameters for tuning.

6.6 Deep Belief Networks (DBNs)

A Deep Belief Network (DBN) is a type of deep neural network (DNN) or a generative graphical model. The DBN is constructed with multiple layers of "hidden units" which are latent variables with the connections across the levels and not in between the units in each layer. DBN can be learned to probabilistically recreate the inputs without supervision when these are trained on the set of example datasets. Next, the DBN's layers act as a feature for detection. Additionally, a DBN can be trained under supervision to categorize the text or object after completing the learning stage. To develop a Novel Q-ADBN for handwritten digit/character recognition, ADAE and Q-learning algorithms are introduced in Deep Belief Network **[4].**

6.7 Restricted Boltzmann Machines (RBMs)

The Adaptive Deep Auto-encoder (ADAE) has been made up of numerous successively stacked RBMs, where the output of one RBM acts as the input of the next. Each RBN is treated as an encoder. The encoder uses unique code to recognise the text .ADAE's hierarchical feature extraction technique is comparable to that of a human's brain **[4].**

6.8 Autoencoders

Autoencoders are the combination of Encoders and Decoders. It is mainly used to learn the compressed representations of the datasets. The Autoencoders should be trained in order to learn the fixed dimensional latent space representation of the given image, making them ideal for the feature extraction **[6].** The output of the autoencoder network is the reconstruction of the input data which is more efficient.

Related Works

Reference	Proposed Algorithm	Purpose	Dataset	Accuracy
1	Convolutional Neural Network	extract feature from raw images	JEITA-HP database Firemaker and IAM database	99.97% 91.81%
2	Codebook model, Clustering, Bayesian classifier, Moore's algorithm	Feature generation and Feature selection, Classifiers using feature vectors	IAM dataset AUT-FH dataset	93.7% 96.9%
3	Convolutional Neural Network + XGradientBoost	CNN - feature extraction XGBoost - Recognition and classification	HECR (CNN+XGBoost)	99.84%
4	Q-ADBN ADAE	extract features from original images using ADAE	MNIST dataset	99.18%
5	Convolutional Neural Network, Long Short Term Memory network	feature extraction Increase the memory of RNN	IFN/ENIT dataset	83%
6	Convolutional Neural Network autoencoder + Support Vector Machine	Feature extractor Classify the images	4600 MODI characters	99.3%
7	Convolutional Neural Network, Algebraic fusion of multiple classifier	Feature extractor Multi-level fusion classifier	MNIST dataset	98%
8	Artificial Neural Network	Classify and Recognize the text from	USTB-Vid TEXT	85%

		images	dataset	
9	Artificial Neural Network	Pre-processingSegmentation, Feature Extraction		95%
10	Convolutional Neural Network Recurrent Neural Network Multidimensional Long-Short Term Memory Networks	Feature Extractor Normalization and standardisation	UPTI dataset	98.12%
11	Convolutional Neural Network	Feature Extractor	NMIST dataset	Testing Acc - 98.85% Training Acc - 98.60%
12	Convolutional Neural Network	Extract the image	NIST dataset	86%
13	Lexicon Convolutional Neural Network, Recurrent Neural Network	Detect common words Extract and classify the images	IAM dataset	99%
14	Convolutional Neural Network	recognize handwritten digits, characters	'hpl-tamil-iso-char' dataset	Training accuracy - 95.16% Testing accuracy - 97.7%
15	Deep Convolutional Neural Network	to extract features from raw data	CMATERdb dataset	Digits - 99.13% Alphabets - 98.31% Characters - 98.18%
16	Convolutional Neural Network	recognize handwritten digits, characters	BanglaLekha-Isolated CMATERdb ISI dataset Mixed dataset	95.71% 98% 96.81% 96.40%
17	Deep Convolutional Neural Network	to extract features from raw data	Ancient Kanada documents	92%
18	Support Vector Machine, Linear Regression, K Nearest Neighbour, Random Forest, MultiNomial Bayes classifier	Classification of images Classification and regression Classify discrete features	IMDB SPAM dataset	85.8% 98.5%
19	Recurrent Neural Network	extract information	IAM database	91.70%
20	Convolutional Neural Network, Bi-Long Short Term Memory networks	extract local features extract global features	Chinese Wikipedia datasets	78%

21	Deep learning technology	Image Preprocessing, Symbol detection, Text recognition	P&ID symbol dataset	97.89%
22	Linear Regression, Long Short Term Memory networks, MultiLayer Perceptron, Decision Tree	Extract features Classify and classify the images	Twitter and Non-Twitter datasets	LR- 99.80% LSVM- 99.78% MLP- 99.12% DT - 99.74%
23	Support Vector Machine	Sign recognition	PSL dataset	80-90%
24	K Nearest Neighbour, Random Forest, Naive Bayes, Support Vector Machine	Text filtering	Protein dataset	KNN - 98.6% RF - 98.5% NB- 96.42% SVM- 97.38%
25	Artificial Neural Network Random Forest, K Nearest Neighbor,	Pattern recognition, Image classification Clustering	Sensor-based SL dataset	ANN -99% RF - 99% K-NN - 98.5%
26	Extreme learning machine (Single hidden layer feedforward neural network)	Uniform random initialization, Xavier initialization,ReLU initialization, Orthogonal initialization	ISI-kolkata Odai numerical, ISI-kolkata Bangla numerical NIT-RKL Bangla Numerical	96.65%, 96.65% 96.89% 97.75%
27	CAPTCHA Convolutional Neural Network	Breaking process and framework	CASIA-HWDB dataset	99.84%
28	Extreme Deep Convolutional Neural Networks, Deep Neural Networks, Linear Regression	Regularization , Normalization and Binarization	SDH2019.2 , MNIST dataset	98.85%
29	Supervised Machine Learning, Support Vector Machine	Preprocessing and feature extraction from images	KVIS Thai OCR Dataset	74.32%
30	Convolutional Neural Network	Character recognition and feature extraction	Devanagari handwritten character dataset	93.73%
31	Convolutional Neural Network	Pre-processing, Character segmentation, Recognition	Devanagari handwritten character dataset	98.47%
32	SEG-WI	Normalization, max-pooling	IAM CVL IFN/ENIT Devanagari	97.27% 99.35% 98.24% 87.24%
33	Beta-elliptic model, Codebook implementation model	Feature selection, feature extraction feature generation	IFN/ENIT dataset	90.02%

34	Sliding Convolutional Neural Network, Slice Convolutional Neural Network	Dataset Collection and Annotation Text recognition	ShopSign dataset	85%
35	Faster R Convolutional Neural Network , RRecurrent Neural Network, Tree Shaped Deep Neural Network	Dataset Preparation for Detection, Custom Feature, Evaluation Prediction, Ligature Recognition	SDAi dataset	95.20%
36	Novel Hybrid network, BoF framework, HMMs, K-means clustering	Pre-processing, Feature Extraction, Classification, Overlapping	P-KHATT dataset	99.95%
37	Encoder- Decoder, Convolutional Neural Network, Bi-Long Short Term Memory networks	Decoding mechanism Feature extraction Sequence analysis	SVT IIIT5K IC03 IC13	84.5% 85.4% 91.9% 91%
38	C Recurrent Neural Network, deep Bi-Recurrent Neural Network, Connectionist Temporal Classification	Fine-tune the Bi-Long Short Term Memory networks, Sequence Labelling, Transcription, Network Training	IIIT5K SVT IC03 IC13	81.2% 82.7% 91.9% 89.6%
39	Deep Convolutional Neural Network	Pre-processing, Character classification Training from scratch, Feature extractor, Fine-tune the CNN	OIHACDB, AHCD	98.86% 99.98%
40	R Convolutional Neural Network ATR-Deep Convolutional Neural Network	Relaxation Convolution, Alternate Training	MNIST dataset	ATR-CNN - Error rate : 0.254±0.014

Table 2 Methods used for text detection and recognition (A Literature Review)

VII. CONCLUSION AND FUTURE SCOPE

Deep learning has a great learning ability and can only benefit from the conclusion, scalar transformation, and background switches. In recent years, deep learning-based detection techniques have become a popular study topic. This study presents a comprehensive overview of deep learning-based detection and recognition strategies that may tackle a variety of sub-problems, such as occlusion, clustering, and lower resolution, using multiple Deep Neural Network methodology (DNNs). The review in this study is based on text detection from handwritten characters using CNN architecture. The study then goes through object detection, face detection, and other types of detection. This review can be applied to advanced discoveries in neural networks and other comparable systems that use deep learning for any detection and classification , and it provides useful and crucial guidelines for future improvement. In the future, Optical Character Recognition will play a vital role to find a way to digitise the words and numbers in physically written text and characters in different languages.

REFERENCES

1. Abdi, M.N., Khemakhem, M., 2015. A model-based approach to offline text-independent Arabic writer identification and verification. Pattern Recognition 48, 1890–1903.

2. AlJarrah, M.N., Zyout, M.M., Duwairi, R., 2021. Arabic Handwritten Characters Recognition Using Convolutional Neural Network, in: 2021 12th International Conference on Information and Communication Systems (ICICS). IEEE.

3. Alom, M.Z., Sidike, P., Hasan, M., Taha, T.M., Asari, V.K., 2018. Handwritten Bangla Character Recognition Using the State-of-the-Art Deep Convolutional Neural Networks. Computational Intelligence and Neuroscience 2018, 1–13. https://doi.org/10.1155/2018/6747098.

4. Alwajih, F., Badr, E., Abdou, S., 2022. Writer adaptation for E2E Arabic online handwriting recognition via adversarial multi task learning. Egyptian Informatics Journal. https://doi.org/10.1016/j.eij.2022.02.007

5. Arafat, S.Y., Iqbal, M.J., 2020. Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning. IEEE Access 8, 96787–96803. https://doi.org/10.1109/access.2020.2994214

6. Boufenar, C., Batouche, M., 2017. Investigation on deep learning for off-line handwritten Arabic Character Recognition using Theano research platform, in: 2017 Intelligent Systems and Computer Vision (ISCV). IEEE.

7. Breuel, T.M., n.d. Handwritten character recognition using neural networks, in: Handbook of Neural Computation. IOP Publishing Ltd.

8. Chowanda, A., Sutoyo, R., Meiliana, Tanachutiwat, S., 2021. Exploring Text-based Emotions Recognition Machine Learning Techniques on Social Media Conversation. Procedia Computer Science 179, 821–828. https://doi.org/10.1016/j.procs.2021.01.099.

9. Elkhayati, M., Elkettani, Y., 2022. UnCNN: A New Directed CNN Model for Isolated Arabic Handwritten Character Recognition. Arabian Journal for Science andEngineering. https://doi.org/10.1007/s13369-022-06652-5.

10. Ghiasi, G., Safabakhsh, R., 2013. Offline text-independent writer identification using codebook and efficient code extraction methods. Image and Vision Computing 31, 379–391. https://doi.org/10.1016/j.imavis.2013.03.002.

11. Guo, H., Liu, Y., Yang, D., Zhao, J., 2021. Offline handwritten Tai Le character recognition using ensemble deep learning. The Visual Computer. https://doi.org/10.1007/s00371-021-02230-2.

12. Han, C., n.d. Neural Network Based Off-line Handwritten Text Recognition System. Florida International University.

13. Hassan, S., Irfan, A., Mirza, A., Siddiqi, I., 2019. Cursive Handwritten Text Recognition using Bi-Directional LSTMs: A Case Study on Urdu Handwriting, in: 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML). IEEE.

14. Hassan, S.U., Ahamed, J., Ahmad, K., 2022. Analytics of machine learning-based algorithms for text classification. Sustainable Operations and Computers 3, 238–248. https://doi.org/10.1016/j.susoc.2022.03.001.

15. Joseph, F.J.J., 2019. Effect of supervised learning methodologies in offline handwritten Thai character recognition. International Journal of Information Technology 12, 57–64. https://doi.org/10.1007/s41870-019-00366-y.

16. Joseph, S., George, J., 2020. Handwritten Character Recognition of MODI Script using Convolutional Neural Network Based Feature Extraction Method and Support Vector Machine Classifier, in: 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP). IEEE.

17. Kavitha, B.R., Srimathi, C., 2022. Benchmarking on offline Handwritten Tamil Character Recognition using convolutional neural networks. Journal of King Saud University - Computer and Information Sciences 34, 1183–1190. https://doi.org/10.1016/j.jksuci.2019.06.004.

18. Kim, C.-M., Hong, E.J., Chung, K., Park, R.C., 2020. Line-segment Feature Analysis Algorithm Using Input Dimensionality Reduction for Handwritten Text Recognition. Applied Sciences 10, 6904. https://doi.org/10.3390/app10196904.

19. Kim, H., Lee, W., Kim, M., Moon, Y., Lee, T., Cho, M., Mun, D., 2021. Deep-learning-based recognition of symbols and texts at an industrially applicable level from images of high-density piping and instrumentation diagrams. Expert Systems with Applications 183, 115337. https://doi.org/10.1016/j.eswa.2021.115337.

20. Kumar, P., Sharma, A., 2020. Segmentation-free writer identification based on convolutional neural network. Computers & Electrical Engineering 85, 106707. https://doi.org/10.1016/j.compeleceng.2020.106707.

21. Li, Z., Teng, N., Jin, M., Lu, H., 2018. Building efficient CNN architecture for offline handwritten Chinese character recognition. International Journal on Document Analysis and Recognition (IJDAR) 21, 233–240. https://doi.org/10.1007/s10032-018-0311-4.

22. Mirza, A., Zeshan, O., Atif, M., Siddiqi, I., 2020. Detection and recognition of cursive text from video frames. EURASIP Journal on Image and Video Processing 2020. https://doi.org/10.1186/s13640-020-00523-5.

23. Narang, S.R., Kumar, M., Jindal, M.K., 2021. DeepNetDevanagari: a deep learning model for Devanagari ancient character recognition. Multimedia Tools and Applications 80, 20671–20686. https://doi.org/10.1007/s11042-021-10775-6.

24. Naz, S., Umar, A.I., Ahmad, R., Siddiqi, I., Ahmed, S.B., Razzak, M.I., Shafait, F., 2017. Urdu Nastaliq recognition using convolutional–recursive deep learning. Neurocomputing 243, 80–87. https://doi.org/10.1016/j.neucom.2017.02.081.

25. Nguyen, H.T., Nguyen, C.T., Ino, T., Indurkhya, B., Nakagawa, M., 2019. Text-independent writer identification using convolutional neural network. Pattern Recognition Letters 121, 104–112. https://doi.org/10.1016/j.patrec.2018.07.022.

26. Pande, S.D., Jadhav, P.P., Joshi, R., Sawant, A.D., Muddebihalkar, V., Rathod, S., Gurav, M.N., Das, S., 2022. Digitization of handwritten Devanagari text using CNN transfer learning – A better customer service support. Neuroscience Informatics 2, 100016. https://doi.org/10.1016/j.neuri.2021.100016.

27. Ptucha, R., Petroski Such, F., Pillai, S., Brockler, F., Singh, V., Hutkowski, P., 2019. Intelligent character recognition using fully convolutional neural networks. Pattern Recognition 88, 604–613. https://doi.org/10.1016/j.patcog.2018.12.017.

28. Qiao, J., Wang, G., Li, W., Chen, M., 2018. An adaptive deep Q-learning strategy for handwritten digit recognition. Neural Networks 107, 61–71. https://doi.org/10.1016/j.neunet.2018.02.010.

29. Rabby, A.S.A., Haque, S., Islam, S., Abujar, S., Hossain, S.A., 2018. BornoNet: Bangla Handwritten Characters Recognition Using Convolutional Neural Network. Procedia Computer Science 143, 528–535. https://doi.org/10.1016/j.procs.2018.10.426.

30. Rahal, N., Tounsi, M., Hussain, A., Alimi, A.M., 2021. Deep Sparse Auto-Encoder Features Learning for Arabic Text Recognition. IEEE Access 9, 18569–18584. https://doi.org/10.1109/access.2021.3053618.

31. Shi, B., Bai, X., Yao, C., 2017. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 2298–2304. https://doi.org/10.1109/tpami.2016.2646371.

32. Shivakumara, P., Sreedhar, R.P., Phan, T.Q., Lu, S., Tan, C.L., 2012. Multioriented Video Scene Text Detection Through Bayesian Classification and Boundary Growing. IEEE Transactions on Circuits and Systems for Video Technology 22, 1227–1235. https://doi.org/10.1109/tcsvt.2012.2198129.

33. Shobha Rani, N., Chandan, N., Sajan Jain, A., R. Kiran, H., 2018. Deformed character recognition using convolutional neural networks. International Journal of Engineering & Technology 7, 1599. https://doi.org/10.14419/ijet.v7i3.14053.

34. Wang, T., Xie, Z., Li, Z., Jin, L., Chen, X., 2019. Radical aggregation network for few-shot offline handwritten Chinese character recognition. Pattern Recognition Letters 125, 821–827. https://doi.org/10.1016/j.patrec.2019.08.005.

35. Wang, Y., Lian, Z., Tang, Y., Xiao, J., 2019. Boosting scene character recognition by learning canonical forms of glyphs. International Journal on Document Analysis and Recognition (IJDAR) 22, 209–219.https://doi.org/10.1007/s10032-019-00326-z

36. Wang, Z.-R., Du, J., 2022. Fast writer adaptation with style extractor network for handwritten text recognition. Neural Networks 147, 42–52. https://doi.org/10.1016/j.neunet.2021.12.002.

37. Wang, Z.-R., Du, J., 2021. Joint architecture and knowledge distillation in CNN for Chinese text recognition. Pattern Recognition 111, 107722. https://doi.org/10.1016/j.patcog.2020.107722.

38. Weldegebriel, H.T., Liu, H., Haq, A.U., Bugingo, E., Zhang, D., 2020. A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters. IEEE Access 8, 17804–17818. https://doi.org/10.1109/access.2019.2960161.

39. Yan, C., Xie, H., Liu, S., Yin, J., Zhang, Y., Dai, Q., 2018. Effective Uyghur Language Text Detection in Complex Background Images for Traffic Prompt Identification. IEEE Transactions on Intelligent Transportation Systems 19, 220–229. https://doi.org/10.1109/tits.2017.2749977.

40. Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu, 2011. A Hybrid Approach to Detect and Localize Texts in Natural Scene Images. IEEE Transactions on Image Processing 20, 800–813. https://doi.org/10.1109/tip.2010.2070803.

41. Zhang, C., Ding, W., Peng, G., Fu, F., Wang, W., 2021. Street View Text Recognition With Deep Learning for Urban Scene Understanding in Intelligent Transportation Systems. IEEE Transactions on Intelligent Transportation Systems 22,4727–4743.https://doi.org/10.1109/tits.2020.3017632.

42. Zhang, X., Liu, X., Sarkodie-Gyan, T., Li, Z., 2021. Development of a character CAPTCHA recognition system for the visually impaired community using deep learning. Machine Vision and Application. https://doi.org/10.1007/s00138-020-01160-8.

43. Zhang, Y., Liang, S., Nie, S., Liu, W., Peng, S., 2018. Robust offline handwritten character recognition through exploring writer-independent features under the guidance of printed data. Pattern Recognition Letters 106, 20–26. https://doi.org/10.1016/j.patrec.2018.02.006.

44. Zhao, H., Liu, H., 2019. Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition. Granular Computing 5, 411–418. https://doi.org/10.1007/s41066-019-00158-6.

45. Zuo, L.-Q., Sun, H.-M., Mao, Q.-C., Qi, R., Jia, R.-S., 2019. Natural Scene Text Recognition Based on Encoder-Decoder Framework. IEEE Access 7, 62616–62623. https://doi.org/10.1109/access.2019.2916616

46. Bora, M. B., Daimary, D., Amitab, K., & Kandar, D. (2020). Handwritten Character Recognition from Images using CNN-ECOC. *Procedia Computer Science*, *167*, 2403–2409. https://doi.org/10.1016/j.procs.2020.03.293

47. Chaturvedi, S., Titre, R.N., Sondhiya, N.R., Khurshid, A. and Dorle, S., 2014. Digits and a special character recognition system using ann and snn models. *International journal of digital image processing*, *6*(06).

48. Gohil, G., Teraiya, R. and Goyani, M., 2012. Chain code and holistic features based OCR system for printed devanagari script using ANN and SVM. *International Journal of Artificial Intelligence & Applications*, *3*(1), p.95.

49. Al-Boeridi, O.N., Syed Ahmad, S.M. and Koh, S.P., 2015. A scalable hybrid decision system (HDS) for Roman word recognition using ANN SVM: study case on Malay word recognition. *Neural Computing and Applications*, *26*(6), pp.1505-1513.

50. Upadhyay, P., Barman, S., Bhattacharyya, D. and Dixit, M., 2011, June. Enhanced Bangla Character Recognition using ANN. In *2011 International Conference on Communication Systems and Network Technologies* (pp. 194-197). IEEE.

51. Hassan, S., Irfan, A., Mirza, A., Siddiqi, I., 2019. Cursive Handwritten Text Recognition using Bi-Directional LSTMs: A Case Study on Urdu Handwriting, in: 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML). IEEE.

52. Joseph, S., George, J., 2020. Handwritten Character Recognition of MODI Script using Convolutional Neural Network Based Feature Extraction Method and Support Vector Machine Classifier, in: 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP). IEEE.

53. Panhwar, M.A., Memon, K.A., Abro, A., Zhongliang, D., Khuhro, S.A., Memon, S., 2019. Signboard Detection and Text Recognition Using Artificial Neural Networks, in: 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC). IEEE.

54. Chowanda, A., Sutoyo, R., Meiliana, Tanachutiwat, S., 2021. Exploring Text-based Emotions Recognition Machine Learning Techniques on Social Media Conversation. Procedia Computer Science 179, 821–828. https://doi.org/10.1016/j.procs.2021.01.099

55. Chiong, R., Budhi, G.S., Dhakal, S., Chiong, F., 2021. A textual-based feature approach for depression detection using machine learning classifiers and social media texts. Computers in Biology and Medicine 135, 104499. https://doi.org/10.1016/j.compbiomed.2021.104499.

56. Wu, C., Fan, W., He, Y., Sun, J., Naoi, S., 2014. Handwritten Character Recognition by Alternately Trained Relaxation Convolutional Neural Network, in: 2014 14th International Conference on Frontiers in Handwriting Recognition. IEEE.

57. Zhang, Y.-K., Zhang, H., Liu, Y.-G., Yang, Q., Liu, C.-L., 2019. Oracle Character Recognition by Nearest Neighbor Classification with Deep Metric Learning, in: 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE.

58. Mirza, A., Zeshan, O., Atif, M., Siddiqi, I., 2020. Detection and recognition of cursive text from video frames. EURASIP Journal on Image and Video Processing 2020. https://doi.org/10.1186/s13640-020-00523-5

59. Shivakumara, P., Phan, T.Q., Lu, S., Tan, C.L., 2013. Gradient Vector Flow and Grouping-Based Method for Arbitrarily Oriented Scene Text Detection in Video Images. IEEE Transactions on Circuits and Systems for Video Technology 23, 1729–1739. https://doi.org/10.1109/tcsvt.2013.2255396

60. Cilia, N.D., De Stefano, C., Fontanella, F., Scotto di Freca, A., 2019. A ranking-based feature selection approach for handwritten character recognition. Pattern Recognition Letters 121, 77–86. https://doi.org/10.1016/j.patrec.2018.04.007

61. Abdulrazzaq, M.B., Saeed, J.N., 2019. A Comparison of Three Classification Algorithms for Handwritten Digit Recognition, in: 2019 International Conference on Advanced Science and Engineering (ICOASE). IEEE.