Text Recognition using Deep Learning: A Review

Shalini Agrahari, Arvind Kumar Tiwari
Department of Computer Science Engineering, Kamla Nehru Institute of Technology Sultanpur, India

Keywords: Deep Learning, Text Recognition, Convolutional Neural Network, Recurrent Neural Network, LSTM.

Abstract: An active field of study, in the domain of text recognition in images, is trying to develop a computer application with the capacity to read the content automatically from images. There is currently a massive demand for the storage of available data on paper records for later use in a machine-readable form. Due to font style, image quality issues, the computer is unable to recognize text. There have been different machine learning and deep learning techniques suggested for text recognition. Current research papers illustrate how different functions boost efficiency and performance in the pattern recognition file. This paper explores and presents the various techniques and to encourage researchers in the field. This paper evaluates and analyzes various techniques for recognizing text from various sources.

1 INTRODUCTION

The text recognition system is a fundamental system used in many applications nowadays. Due to digitalization, there is a huge demand for storing data into the computer by converting documents into digital format. It is difficult to recognize text in various sources like text documents, images, and videos, etc. due to some noise. The text recognition system is a technique by which recognizer recognizes the characters or texts or various symbols. The text recognition system consists of a procedure of transforming input images into machine-understandable format (K. R. Singh, 2015; Polaiah B, 2019; Yogesh Kumar D, 2018). There are two types of recognition: online text recognition and offline text recognition whether online recognition system includes tablet and digital pen, while offline recognition includes printed or handwritten documents. Also, offline text recognition is further divided into handwritten text recognition and printed text recognition. A text recognition procedure is carried out by some phases: first is pre-processing to enhance the quality of input image by doing some operations like noise elimination and normalization etc., second is segmentation to segment the input image into single characters, third is feature extraction to extract important information from the input image by applying different feature extraction techniques such as Histograms, etc., forth is classification or decision-making phase which compares the extracted input feature to the stored pattern, and assigns them into correct character class, and the last one is post-processing which improves the recognition rate by filtering and correcting the output obtained by classification phase (Ayush Purohit, 2016; Manoj Sonkusare, 2016).

Many machine learning algorithms, deep learning algorithms, and datasets are used to detect and recognize text. The idea of developing a predictive model based on experience is involved in machine learning. The deep neural network can model convoluted and non-linear connections and the creation of models. Many researchers are researching this field to find an accurate system for recognizing text.

2 DIFFERENT DEEP LEARNING APPROACHES

Deep learning's objective approach is to solve the complexity of the sophisticated factors of the input through using high–level features. The notion of deep learning technology is that there is nothing that inherently challenges the software to increase performance, e.g. handwriting recognition of the machines achieves human performance levels. There are numerous types of structures and algorithms that are useful in the formulation of the deep learning idea.

Agrahari, S. and Tiwari, A.
Text Recognition using Deep Learning: A Review.
DOI: 10.5220/0010563300003161
In *Proceedings of the 3rd International Conference on Advanced Computing and Software Engineering (ICACSE 2021)*, pages 106-114
ISBN: 978-989-758-544-9

Some of the deep learning's important approaches are:

2.1 Convolutional Neural Network (CNN)

CNN (Edward Grefenstette, 2014) is a specialized type of neural network model which is designed for working with image data. CNN is widely used in image recognition, image classification, object detection and face recognition, etc. In CNN, the input image passes through a series of convolutional layers, pooling layers, fully connected layers, and finally produces an output which can be a simple class or probability of class that best describes the image. CNN can learn multiple layers of feature representations of an image by applying different techniques. In this approach, a computer performs image classification by looking for low-level features such as edges and curves and then building up to a more abstract concept through a series of convolutional layers. CNN provides greater precision and improves performance because of its exclusive characteristics, such as local connectivity and parameter sharing.

2.2 Recurrent Neural Network (RNN)

RNN (Junyoung C., 2014) is useful when it is required to predict the next word of sequence. If we are trying to use such data where sequence matches, we need a network that has access to some prior knowledge about the data. In this approach, output from the previous step is fed as input to the current step. The architecture of RNN includes three layers: input layer, hidden layer, and output layer. The hidden layer remembers information about sequences. RNN model has a memory that remembers all the information about what has been calculated. Application areas of RNN include sequence classification such as sentiment classification and video classification etc.; sequence labeling such as image captioning and named entry recognition etc.; and sequence generation such as machine translation etc. Recurrent Neural Network is useful in time series prediction and it has flexibility for handling various types of data.

2.3 Long Short-Term Memory (LSTM)

LSTM (S. Hochreiter, 1997) overcomes the problems of RNN model. RNN model suffers from short-term memory. RNN model has no control over which part of the information needs to be carried forward and how many parts need to be forgotten. A memory unit called a cell is utilized by the LSTM which can maintain information for a sufficient period. LSTM networks are a special kind of RNN, capable of learning long dependencies. LSTM has a cell state which carries information throughout the processing of the sequence. This model contains interacting layers in a repeating module. Forget gate layer is responsible for what to keep and what to throw from old information. The input modulation gate layer and input layer are responsible for what new information is stored in the cell state. And last output layer gives output based on cell state.

3 RELATED WORK

In literature, there are various researchers have been proposed a deep learning-based approach for text recognition. A multilingual handwriting recognition system has been proposed using gated convolutional recurrent neural networks (Blutche T., 2017). In this model, a convolutional encoder, an interface that converts 2D images into 1D representation, and a Bi-LSTM layer were used. The given architecture was faster than multi-dimensional-LSTM for text recognition. On IAM dataset, the character error rate (CER) was improved by 3.2% on line level and 3.3% on the paragraph level. (Reeve Ingle, 2019) have described a model for online handwriting data recognition, named "A Scalable Handwritten Text Recognition System". Gated recurrent convolutional layers (GRCLs) were used in this model. GRCL blocks are useful for large amounts of data. Historical documents and IAM dataset were used for training data. This model improved character error rate (CER) by 4.0 to 12.4% and word error rate (WER) by 10.8 to 30.1%. (Chung J., 2019) have introduced a framework to recognize offline handwritten text. Text localization and text recognition processes were used to localize handwritten test and converted images of words into strings. In this model, CNN-biLSTM network was implemented. For evaluation, IAM dataset was used in which CER improved by 6.4 to 28.3%. Computational cost was reduced, if compared to previous methods. (Carbonell M., 2018) has proposed Joint recognition of handwritten text and named entities with a neural end-to-end model. In this model, CNN, BLSTM, and CTC model were combined together for performing handwritten text recognition (HTR). 4 convolutional layers and 3 stacked BLSTM layers were used in this architecture. Marriage records were used as a dataset. (Bluche, 2016) has proposed a model which is a modification of Multi-Dimensional Long Short-Term Memory Recurrent Neural Networks (MDLSTM-RNNs), for processing of handwritten paragraphs without an explicit line segmentation. IAM database was used as a dataset. CER improved by 6.8 to 8.4%. The CTC error at the paragraph level is minimized by line segmentation. (Moysset, 2019) have discussed Manifold Mixup improves text recognition with the CTC loss model. Manifold Mixup is a data augmentation method that features maps to images. This model performed experiments on the Maurdor dataset (french) and IAM dataset. 9.39% (without mixup) and 8.9% (with mixup) improvement were achieved on maurdor dataset, and on IAM dataset, 4.68% (without mixup) and 4.64% (with mixup) improvement were achieved. This model significantly enhanced text recognition results on various sets of handwritten data of different sizes and languages. (Wenniger, 2019) have proposed a model named, No Padding Please: Efficient Neural Handwriting Recognition Model. For neural handwriting recognition (NHR), efficient MDLSTM based models have been developed. This example-packing proposed method replaced stacking of waste padded examples with efficient tiling in a 2-dimensional grid. For evaluation, IAM dataset was used. Speed improvement was achieved in this model. (Louradour, 2014) have introduced an active curriculum learning model, which was used to recognize the text for localization and classification. In this paper, stochastic gradient descent was easily accelerated and short sequences were recognized before all training sequences were trained. IAM, RIMES, and OpenHaRT were used for experiments. CER% was dropped from 22% to 17% on IAM dataset. A slight improvement has been achieved in both RIMES and OpenHaRT datasets.

(Poulos, 2019) have proposed a model i.e. called Character-Based Handwritten Text Transcription with Attention Networks. An attention-based encoder-decoder network was used to train to handle sequences of characters rather than words for handwritten text transcription. In this model, BLSTM as an encoder and gated recurrent unit (GRU) as a decoder were used. This approach achieved the lowest test error and outperformance than other RNN based models. For evaluation, IAM databaset was used as dataset. CER = 16.9% achieved in this model with softmax attention. (Aradillas, 2018) have described an approach to enhance the recognition of handwriting texts in small databases with neural networks. CNN, BLSTM, and CTC algorithms were used to reduce training data for handling offline handwritten text recognition. The transfer learning approach was used for learning parameters form a bigger database. IAM dataset, Washington and Parzinal were used as datasets; CER=3.0% achieved on IAM dataset, and this model obtained good results from Washington and Parzinal database. (Chowdhury, 2018) have proposed a model to recognize the offline handwritten text from images. The Encoder-Decoder network was being used to plot text in the image to a character sequence; BLSTM and LSTM were used as encoder and decoder respectively. Beam Search algorithm was applied for searching the best sequences. IAM and RIMES datasets were used for experimental results. Word level accuracy was achieved by 3.5 % on IAM dataset and 1.1 % on RIMES dataset. (Puigcerver, 2017)have proposed a model in which a multidimensional recurrent layer might not be necessary. Convolutional layer, 1D-LSTM layers, and RmsProp algorithm were used in this method. RmsProp algorithm updated the parameters of the model incrementally. Two widely used datasets- IAM and RIMES dataset were used for experimental results in this method. CER dropped from 8.2% to 6.3% on IAM dataset and 3.3% to 2.6% on RIMES dataset. (Michael, 2019) have introduced a method; attention-based sequence-to-sequence method. It was encoder-decoder-based network in which deep CNN and BLSTM algorithms were used in encoder and LSTM algorithm used in decoder. Here, 6 different types of attention mechanism were tried. Different datasets like IAM Handwriting Database, ICFHR2016 READ data set (Bozen), and StAZH data set was used in this method. It achieved an average CER of 4:66% on bozen dataset and CER of 4:87% on IAM dataset. (Lei Tang, 2009) have suggested the MetaLabeler model to automatically determine the appropriate dataset. MetaLabeler model has divided into three versions: score-based, content-based, and ranked-based. Content-based MetaLabeler outperformed than other methods. With MetaLabeler, both measures micro-f1 and macro-f1 improved by 1-9% at a different level concerning hierarchical models. (Alex Graves, 2014) has introduced an LSTM-RNN based method that generates complex, real and discrete sequences with long-range structures. Prediction network was for predicting text; this model was also used for recognizing online handwritten text. IAM database was used as a dataset to define the character sequences. LSTM and RNN algorithms were used in this model. This system generated highly realistic cursive handwriting in a wide variety. (Vu Pham,

2014) have proposed a model in which DropOut and DropConnect methods are used. LSTM and RNN algorithms were also used in this method for text recognition. Three handwriting datasets were used for evaluation Rimes, IAM, and open heart. DropOut improved error rate always.

(Khaoula, 2011) have introduced a model in which text can be detected and recognized in digital videos. A full Optical Character Recognition (OCR) System was developed for text recognition in videos. The segmentation method has been applied to segment images into single characters and to recognize them. This OCR-based system outperformed in style and size variabilities, background complexity, and indexing multimedia videos. French TV news videos database was used as a dataset in this model and CER improved by 95%. (Voigtlaender, 2016) have introduced GPU based model that lowers training times. CUDA, cuBLAS, and MDLSTM were directly used to implement this model. For result evaluation, IAM and RIMES datasets were used. We achieved a WER of 7.1 % on IAM dataset. (Doetsch, 2016) have introduced a method to recognize offline handwritten text. The Proposed model was a bidirectional decoder network which was superior to a unidirectional decoder network. A Recurrent encoder, bidirectional decoder, and content-based attention mechanism were combined for consistent improvements. RIMES database was used as a dataset for experimental results. This model achieved WER of 3.6% CER of 1.3%. (Zhuoyao, 2015) have given a model which used GoogLeNet and Directional Feature Maps for high performance in HCCR. A deep architecture has been used for high performance; in which multiple layers were used to implement HCCR-GoogLeNet. In this model, ICDAR 2013 dataset was used as an offline HCCR competition dataset. HCCR-GoogLeNet achieved a recognition accuracy of 96.35%. (Baoguang, 2016) have proposed a technique to recognize scene text in image-based sequence recognition. The network architecture of this model was based on CRNN which is a combination of DCNN and RNN algorithms. Different datasets were used for performance evaluation. By the result, given by the CRNN based model, made this proposed model compact and efficient.

(Zheng Zhang, 2016)have described a model to detect text in natural images. For text blocks detection, Text-Block FCN was used. Character-Centroid FCN eliminated false text line candidates. Two multi-oriented text datasets- MSRATD500 and ICDAR2015; and one horizontal text dataset - ICDAR2013; were used for result evaluation. In the handling of multi-oriented text, this model usually accomplished the state of art efficiency but did not achieve a perfect performance. (Swapnil, 2016) have introduced a model to recognize Hindi handwritten words based on HMM and symbol tree. Segmentation and classification method were used for text recognition. In this model, the symbol tree has been created by possible sequences of recognized text. HMM was used for robust classification. Akshara's Segmentation is critical work to generate accuracy. This method gives 89% accuracy on 10,000 words. (Saumya, 2014) have described a method to Multi-script Identication from Printed Words. For script recognition at the word level HoG was used. LPG was used to capture image texture. Both HoG and LPG-based classification were used as feature descriptors for the word image. MILE database was used as a dataset in this model. This model gives 97.7% accuracy on MILE dataset. (Hung T. N, 2019) have proposed a model for text-independent writer identification. The CNN-based approach was used to extract local and global features. To create a high volume of data, a random sampling method was also used. JEITA-HP database and IAM database were used in this model as datasets. This method achieved a 99.97% identification rate on JEITA-HP database, and also obtained 91.5% accuracy higher than handcrafted features. (Nam-Tuan Ly, 2017) have introduced a method to recognize Offline Handwritten Japanese Text. DCRN approach was applied which was a combination of CNN, BLSTM, and CTC decoder algorithms; CNN for feature extractor, BLSTM at recurrent layer, and CTC decoder for translating prediction into sequences. TUAT Kondate database (a Japanese texts, images, etc. database) was used as a dataset. This DCRN based model consistently outperformed than other segmentation-model in both sequence error rate and label error rate. (Weixin Yang, 2015) have introduced an approach to recognize Chinese characters using domain-specific information. The proposed method was generally an enhancement of the DCNN approach. This method included non-linear normalization, deformation, imaginary strokes, and path signature, etc. Hybrid serial-parallel (HSP) strategy with DCNN significantly improved state-of-the-art. In this method, accuracy was achieved by 97.20% on the CASIA-OLHWDB1.0 dataset and 96.87% on CASIAOLHWDB1.1. (Irfan Ahmad, 2014) have described their model for Arabic Text Recognition. This proposed model was based on sub-character HMM models. This recognizer allowed to share common patterns and Arabic characters between different shapes and different characters respectively. This sub-character HMM model included space and connector model; where space model provided flexibility. By this method, a compact, efficient and robust model has been developed. IFN/ENIT database was used for evaluation and it achieved recognition rates of 85.12%. (Gernot, 2016) have introduced their investigations for text recognition using class-based contextual modeling. There was a problem of inadequate training using the Contextual HMM model for recognition. Three processes data preparation, model training, and decoding were used in the proposed model. IFN/ENIT databases were used as a dataset; which gives better results than standard contextual HMM systems. (Minghui, 2017) have introduced a TextBoxes model for scene text recognition. This textboxes model included text detection, word spotting, and end-to-end recognition steps. SynthText, IC13, and Street View Text datasets were used for text localization performance. This model gives text localization performance with high accuracy and efficiency but there were problems in overexposure and large character spacing. (Baoguang, 2018) have suggested a scene text recognition model based on attention with flexible rectification. This technique included a rectification network and a recognition network where it was the responsibility of the rectification network to correct text in images, and the recognition network estimated plain text from the corrected image. This proposed model was trained on SynthText, IIIT5k-Words, and Street View Text (SVT) datasets. This model addressed the issues of irregular text recognition problems and gave greater efficiency in the recognition of cropped text.

(Xiang Bai, 2016) have investigated a method for script identification that finds script in natural images. Discriminative clustering has been applied; by which this model was called Discriminative Convolutional Neural Network (DisCNN) system. For validation, a SIW-13 dataset was evaluated. This technique did not include processes such as binarization, segmentation, or hand-crafted characteristics. This method effectively performed script identication in images, videos, and documents. (Xiang Bai,2017) have introduced a model for classifying images whether it contains text or not. A CNN-based model variant called Multi-scale Spatial Partition Network (MSP-Net) has been developed. This process categorized images very efficiently in a single forward propagation by foreseeing all sections at once. This method included TextDis benchmark database, the ICDAR2003 database, and Hua's database as datasets for evaluation. This text/non-text image classification method gives effective results compared with other methods. (Sheng Zhang, 2018) have introduced a new Feature Enhancement Network for accurate scene text detection. Text Detection Refinement algorithm was used for refining text. In this model, FE-RPN enhanced text features, and Hyper Feature Generation module was used for text detection refinement. Two ICDAR 2011 and ICDAR 2013 datasets were used for proving the effectiveness of this approach. The state-of-the-art outcomes are ultimately achieved by this technique. (Xiaoxue Chen, 2020) have proposed an Adaptive embedding gate (AEG) module to address the problems of improper use of previous predictions in the scene text recognition attention decoder. By using character language modeling, the AEG module adaptively strengthens the current prediction in the decoding stage. Two components: a convolutional encoder and recurrent attention-based decoder network were used to generate target sequences. Several datasets were used for result evaluation. The efficiency of state-of-the-art outcomes is consistently improved by this AEG module. (Shangbang Long, 2018) have presented a framework for arbitrary shapes text detection. The proposed method TextSnake tackled the problem generated from free-form text instances. A fully Convolutional Network (FCN) algorithm was used for attribute estimation. This method gives a flexible representation of arbitrary shapes text instances. SynthText was used for the pre-training network. TotalText, CTW1500, and MSRA-TD500 datasets were used for achieving result performance. Text detection-based TextSnake outperformed on TotalText dataset. (Christian Bartz, 2017) have introduced a model for language identification. Automatic Speech Recognition (ASR) systems detected languages but this proposed model solved the problem of images, not the audio domain. This model used a hybrid algorithm based on CRNN, a combination of CNN and RNN. YouTube News dataset and EU Speech Repository etc. were used as datasets whether datasets are split into training, validation, and testing sets. This algorithm recognized a wide variety of languages.

(Shangbang Long, 2020) have introduced a UnrealText which renders realistic images via a 3D graphics engine. This engine showed real appearances of both text and scene both. The viewfinder module was used for exploring the camera's location, viewpoint, and rotation from 3D scenes. Based on UE4.22 and the UnrealCV plugin, this proposed engine has been developed. This method adopted the techniques of ASTER model. Synthetic datasets were used for training such as Synth90K and SynthText. To detect and recognize texts, this method provided effectiveness. (Christian Bartz, 2018) have described a deep neural network-based model called, SEE. It is a semi-supervised method of detection and recognition of end-to-end text. For text detection, a spatial transformer is used which is a combination of localization network, grid generator, and differentiable interpolation method. SVHN dataset and FSNS dataset are used for showing the performance of the proposed method. In-text detection and recognition, this paper outperforms but is still unable to identify text in irregular places. (Joseph Bethge, 2019) have mentioned a strategy for scene text detection which consisted of off-the-shelf building blocks for neural networks. There were two parts to this network system, one was the localization network and the other was the recognition network. Localization included ResNet based feature extractor and spatial transformer, and recognition network included ResNet based feature extractor and transformer. Datasets such asICDAR 2013(IC13), IIIT5K, SynthText, SVTP, and CUTE80 were used for result evaluations. In comparison with other approaches, this proposed work outperformed the state results. (Maroua Tounsi, 2016) has introduced a model for scene character recognition(SCR). The proposed model was based on a Bag of Features (BoF) model that uses supervised learning dictionaries. This method also used the strategy of a sparse neural network model for character recognition. Chars74K, ICDAR2003, and ARASTI datasets were used for proving this technique. The experimental results demonstrated the performance of the method suggested.

(Ikram Moalla, 2016) have been suggested as a technique to recognize text in captured images by camera. This model extended the features of BoF based model using deep learning for SCR. To improve recognizer performance, a deep Sparse Auto-encoder (SAE)-based strategy was also adapted. Datasets such as Chars74K, ICDAR2003, and ARASTI were used for evaluation. This deep learning-based architecture gives better representation and better recognition accuracy. (Fei Yin, 2017) have described sliding convolutional character models for STR. This method uses a strategy of a Convolutional feature map. CTC algorithm has been applied for normalizing and decoding character outputs on sliding windows. This proposed method avoided character segmentation difficulty and gradient exploding. Different datasets such as IIIT-5K, SVT, ICDAR03/13, and TRW1 were taken in this technique to show the performance of the proposed method. The recognizer was fast and provided state-of-the-art techniques with a competitive advantage. (Yi-Chao Wu, 2018) have suggested an approach to do text recognition similar to the human reading mechanism. The process of SCAN recognizer was like the movement of the human eye during the reading text. This recognizer included a sliding window layer, a convolutional feature extractor, and a convolutional encoder-decoder. This proposed method was evaluated on IIIT5k, SVT, and ICDAR 2003/2013 datasets. This recognition method gives high interpretation and high performance. (Praveen Krishnan, 2016) have described a deep convolutional feature representation model for spotting words and recognizing texts. HWNet architecture was used for word spotting. It was a CNN-based network, with concepts of query-by-string and query-by-example. IAM Handwriting Database was used as a dataset for result evaluations. Word error rate (WER) improved by 6.69% and character error rate(CER) improved by 3.72% in the proposed method. (Kartik Dutta, 2018) have investigated an End2End embedding framework for text recognition and word-spotting in text documents and images. HWNet embedding architecture and CRNN algorithm were used for word-spotting and word-recognition respectively in this method. IAM handwritten dataset was used for experimental results. Word spotting has been evaluated using mean average precision (mAP) of 0.9509 for query-by-string. CER and WER improved by 2.66% and 5.10% respectively on word recognition. (Minesh Mathew, 2018) have proposed a modified version of CNN-RNN hybrid architecture for handwritten recognition in offline document images. Different procedures like synthetic data for pre-training, image normalization method, and data transformation and distortion were jointly used for recognition in this technique. Two modern datasets IAM and RIMES, and one historical dataset, George Washington (GW) dataset were used for document analysis. This proposed method significantly improved the recognition rate at line level and word level. (CV Jawahar, 2016) have introduced a framework for data generation in documents. A data augmentation scheme is used for rendering synthetic data. Synthetic data is created based on fonts and style for the word images. IAM handwriting dataset and IIIT-HWS dataset are used for experimental results. This framework gives good performance in recognition but does not address problems of the cursive property.

(Siddhant Bansalhave, 2020) given a word recognition method using deep embeddings representation. CNN-RNN hybrid architecture was used to convert the content of images into text forms, and fusion and re-ranking methods were used for word retrieval. In this paper, Hindi documents were used as a dataset to validate this method. Average and max fusion were used, while retrieving, for result improvement. In the mAP, the word recognition rate increased by 1.4 percent and the recovery rate by 11.13 percent.

4 CONCLUSIONS

Text recognition is a challenging task due to different writing styles in different languages. In this paper, we have studied how several steps such as segmentation, feature extraction, and classification have been used in text recognition. Here deep learning approaches have been also discussed which helps recognize text. This paper analyzed and explored in direction of text recognition. The analysis in this paper showed that there is still scope to improve the algorithms, as well as improve the recognition rate of words.

REFERENCES

Alex Graves (2014).Generating Sequences With Recurrent Neural Networks. 1308.0850v5 [cs.NE].

Aradillas, J. C., Murillo-Fuentes, J. J., & Olmos, P. M. (2018). Boosting Handwriting Text Recognition in Small Databases with Transfer Learning. arXiv preprint arXiv:1804.01527.

AyushPurohit and Shardul Singh Chauhan (2016). A Literature Survey on Handwritten Character Recognition", IJCSIT,Vol. 7 (1).

Baoguang S., Mingkun Y., Xinggang W., Pengyuan L., C. Yao, and Xiang B (2018)," ASTER: An Attentional Scene Text Recognizer with Flexible Rectification",IEEE.

Baoguang Shi, Xiang Bai, & Cong Yao (2016).An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition.IEEE (TPAMI).

Baoguang Shi, XiangBai n, &CongYao (2016). Script identification in the wild via discriminative convolutional neural network.Pattern Recognition(vol-52).

Bluche T., & Messina R. (2017, November). Gated convolutional recurrent neural networks for multilingual handwriting recognition. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 646-651). IEEE.

Bluche, T. (2016). Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. In Advances in Neural Information Processing Systems (pp. 838-846).

Carbonell, M., Villegas, M., Fornés, A., &Lladós, J. (2018, April). Joint recognition of handwritten text and named entities with a neural end-to-end model. In 2018 13th IAPR International Workshop on Document Analysis Systems (DAS) (pp. 399-404). IEEE.

Chowdhury, A., &Vig, L. (2018). An efficient end-to-end neural model for handwritten text recognition. arXiv preprint arXiv:1807.07965.

Christian Bartz, Haojin Yang, and Christoph Meinel (2018). SEE: Towards Semi-Supervised End-to-End Scene Text Recognition. Proceedings of the AAAI Conference on Artificial Intelligence.

Christian Bartz, Joseph Bethge, Haojin Yang, and Christoph Meinel (2019). KISS: Keeping It Simple for Scene Text Recognition. arXiv preprint arXiv:1911.08400.

Christian Bartz, Tom Herold, Haojin Yang, Christoph Meinel (2017). Language Identification Using Deep Convolutional Recurrent Neural Networks. International conference on neural information processing, Springer.

Chung, J., &Delteil, T. (2019, September). A Computationally Efficient Pipeline Approach to Full Page Offline Handwritten Text Recognition. In 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW) (Vol. 5, pp. 35-40). IEEE.

Fei Yin, Yi-Chao Wu, Xu-Yao Zhang, and Cheng-Lin Liu (2017).Scene Text Recognition with Sliding Convolutional Character Models.

Hung T. N., Cuong T. N., Takeya Ino, BipinI.,& Masaki N. (2019). Text-Independent Writer Identification using Convolutional Neural Networks.Pattern Recognition Letters, (vol-121).

Ingle, R. R., Fujii, Y., Deselaers, T., Baccash, J., &Popat, A. C. (2019). A Scalable Handwritten Text Recognition System. arXiv preprint arXiv:1904.09150.

Irfan Ahmad, and Gernot A Fink(2016).Class-Based Contextual Modeling for Handwritten Arabic Text Recognition.(ICFHR),IEEE.

Irfan Ahmad, Gernot A. Fink, and Sabri A. Mahmoud(2014).Improvements in Sub-Character HMM Model Based Arabic Text Recognition. International Conference on Frontiers in Handwriting Recognition.IEEE.

Junyoung C., C. Gulcehre, KyungHyun Cho, and YoshuaBengio.(2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

K. R. Singh," A Technical Review on Text Recognition from Images (2015). IEEE Sponsored 9th International

Conference on Intelligent Systems and Control (ISCO).

Kartik Dutta, Praveen Krishnan, Minesh Mathew, and CV Jawahar (2018). Improving CNN-RNN Hybrid Networks for Handwriting Recognition. International Conference on Frontiers in Handwriting Recognition.

Khaoula E., C. Garcia,& Pascale S.(2011).Comprehensive Neural- Based Approach for Text Recognition in Videos using Natural Language Processing.ICMR.

Lei Tang, Suju R., & Vijay K. N.(2009).Large Scale Multi-Label Classification via MetaLabeler.

Louradour, J., &Kermorvant, C. (2014, April). Curriculum learning for handwritten text line recognition. In 2014 11th IAPR International Workshop on Document Analysis Systems (pp. 56-60). IEEE.

Manoj Sonkusare and Narendra Sahu (2016). A Survey On Handwritten Character Recognition (HCR) Techniques For English Alphabets.Advances in Vision Computing: An International Journal (AVC) Vol.3, No.1.

MarouaTounsi, IkramMoalla, Adel M Alimi (2016). Supervised Dictionary Learning in BoF Framework for Scene Character Recognition. 23rd International Conference on Pattern Recognition (ICPR).

MarouaTounsi, IkramMoalla, Frank Lebourgeois, Adel M Alimi(2018). Multilingual Scene Character Recognition System using Sparse Auto-Encoder for Efficient Local Features Representation in Bag of Features.

Michael, J., Labahn, R., Grüning, T., &Zöllner, J. (2019). Evaluating Sequence-to-Sequence Models for Handwritten Text Recognition. arXiv preprint arXiv:1903.07377.

Minghui L., Baoguang S., Xiang Bai, Xinggang W., &WenyuL.(2017), ”TextBoxes: A Fast Text Detector with a Single Deep Neural Network”, AAAI Conference on Artificial Intelligence.

Moysset, B., & Messina, R. (2019). Manifold Mixup improves text recognition with CTC loss. arXiv preprint arXiv:1903.04246.

NalKalchbrenner, Edward Grefenstette, and Phil Blunsom (2014). A convolutional neural network for modelling sentences. In Proceedings of ACL.

Nam-Tuan Ly, Cuong-Tuan Nguyen, Kha-Cong Nguyen, & Masaki N.(2017). Deep Convolutional Recurrent Network for Segmentation-free Offline Handwritten Japanese Text Recognition. IAPR (ICDAR) (vol-7).

P. Doetsch, A. Zeyer, and H. Ney (2016). Bidirectional decoder networks for attention-based end-to-end offline handwriting recognition,” International Conference on Frontiers in Handwriting Recognition, pp. 361–366.

P. Voigtlaender, P. Doetsch, and H. Ney (2016). Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. ICFHR.

Polaiah B., Naga s., Gautham K. P., and S D Lalitha Rao Sharma Polavarapu (2019). Handwritten Text Recognition using Machine Learning Techniques in Application of NLP.(IJITEE) ISSN: 2278-3075, Volume-9 Issue-2.

Poulos, J., & Valle, R.(2019). Character-Based Handwritten Text Transcription with Attention Networks.

Praveen Krishnan, and CV Jawahar (2016).Generating Synthetic Data for Text Recognition.

Praveen Krishnan, Kartik Dutta, and CV Jawahar (2016). Deep eature Embedding for Accurate Recognition and Retrieval of Handwritten Text. International Conference on Frontiers in Handwriting Recognition.

Praveen Krishnan, Kartik Dutta, and CV Jawahar (2018).Word Spotting and Recognition using Deep Embedding. Document Analysis Systems.

Puigcerver, J. (2017). Are multidimensional recurrent layers really necessary for handwritten text recognition? In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 67-72). IEEE.

S. Hochreiter and J. Schmidhuber (1997). Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780..

Saumya J., KapilMehrotra, AtishVaze, &SwapnilBelhe (2014). Multi-script Identication from Printed Words. International Conference Image Analysis and Recognition.

Shangbang Long, and Cong Yao(2020). UnrealText: Synthesizing Realistic Scene Text Images from the UnrealWorld. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(pages. 5488-5497).

Shangbang Long, JiaqiangRuan, Wenjie Zhang, Xin He, Wenhao Wu, and Cong Yao(2018). TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes.

Sheng Zhang, Yuliang Liu, LianwenJin, Canjie Luo (2018). Feature Enhancement Network: A Refined Scene Text Detector. AAAI Conference on Artificial Intelligence(vol-32).

Siddhant Bansal, Praveen Krishnan, and CV Jawahar(2020). Fused Text Recogniser and Deep Embeddings Improve Word Recognition and Retrieval. International Workshop on Document Analysis Systems. Springer, Cham (Pp.309-323).

SwapnilBelhe, Chetan P., Akash D., Saumya J., &KapilM.(2016).Hindi Handwritten Word Recognition using HMM and Symbol Tree. Workshop on Document Analysis and Recognition.

Vu Pham, T. Bluche, Christopher K.,& J. Louradour (2014). Dropout improves Recurrent Neural Networks for Handwriting Recognition. arXiv:1312.4569v2 [cs.CV].

Weixin Y., Lianwen J., ZechengXie, &ZiyongFeng(2015). Improved Deep Convolutional Neural Network For Online Handwritten Chinese Character Recognition using Domain-Specific Knowledge.(ICDAR), IEEE.

Wenniger, G. M. D. B., Schomaker, L., & Way, A. (2019). No Padding Please: Efficient Neural Handwriting Recognition. arXiv preprint arXiv:1902.11208.

Xiang Baia, Baoguang Shia, ChengquanZhanga, Xuan Caib, Li Qib (2017). Text/non-text image classification in the wild with convolutional neural Networks. Pattern Recognition(vol-66).

Xiaoxue Chen, Tianwei Wang, Yuanzhi Zhu, LianwenJin, Canjie Luo (2020). Adaptive embedding gate for attention-based scene text recognition. Neurocomputing(vol-381).

Yi-Chao Wu, Fei Yin, Xu-Yao Zhang, Li Liu, and Cheng-Lin Liu (2018). SCAN: Sliding Convolutional Attention Network for Scene Text Recognition. arXiv preprint arXiv:1806.00578.

Yogesh Kumar D. and Pulkit Jain (2018). Comprehensive Survey on Machine Learning Application for Handwriting Recognition. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 8 pp. 5823-5830.

Zheng Zhang, Chengquan Zhang, Wei Shen, Cong Yao, Wenyu Liu, & Xiang Bai (2016),"Multi-Oriented Text Detection with Fully Convolutional Networks".IEEE.

Zhuoyao Z., Lianwen J., &ZechengXie(2015). High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps. ICDAR ,(pp.846-850).