Global Transitions Proceedings 3 (2022) 230–234

Contents lists available at ScienceDirect

Global Transitions Proceedings

journal homepage: http://www.keaipublishing.com/en/journals/global-transitions-proceedings/

# An improved method for text detection using Adam optimization algorithm

Himani Kohli[a], Jyoti Agarwal[b], Manoj Kumar[c,*]

[a] Department of CSE, Amity University Uttar Pradesh, Noida, India, 201309
[b] Department of CSE, Graphic Era Deemed to be University, Dehradun, India, 248002
[c] School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India, 248007

A R T I C L E I N F O

*Keywords:*
Artificial Intelligence
Python
OpenCV
Optical Character recognition
Text Detection

A B S T R A C T

Optical Character Recognition (OCR) is an automatic identification technique which is applied in different application areas to translate documents or images into analysable and editable data. Printed or typed characters are easy to recognize as they have well defined shape and size, but this is not true in case of handwritten text. Handwriting of every individual is different so OCR face difficulty to recognize the characters. In past, researchers have been used different Machine Learning and Artificial Intelligence tools and techniques to analyse handwritten and printed documents and also worked to create an electronic format file from them. It is difficult to reuse this information as it is very difficult to search the content from these documents by lines or words. To solve this problem, OpenCV technique is used in this research work which focuses on training and testing of neural network model to conduct Document Image Analysis. The proposed model is named as J&M model for Text Detection from Hand written images. Implementation of research work is done in Python on MNIST database of handwritten digits. From this research work, 99.5% of training accuracy and 99% of testing accuracy was achieved along with training loss of 1.5%.

## 1. Introduction

In recent year, text recognition has received a wide popularity and also being used in various areas like auto- reading license plate, sign boards etc. [1]. Printed or typed characters can be easily identified because they have well defined shape and size. But the condition is different for handwritten documents because they are not having a defined shape or size. If a person is writing the same sentence more than one time, there is a possibility that the handwriting may not appear exactly the same [2]. Handwritten characters differ by 12 characteristics such as characters and words, height, etc. [1]. Because of these differences it is not easy to recognize handwritten text. Recognition of handwritten manuscripts is required for identifying valuable information from the text. Optical Character Recognition (OCR) is a system that can be used to detect text from images as well as for detecting the handwritten text information. It converts the input text into machine encoded format [3]. OCR technique enables the user not only in digitizing the handwritten documents, but also create their digital files [4]. This scanned data can be easily maintained which can be used in future whenever required.

OCR system depends on feature extraction and classification. Handwritten OCR is a subfield of OCR which is categorized into online and offline OCR system. First OCR system was developed in 1990 and later advancements are done based on the changing technologies. In last decade, different Machine Learning approaches have been also applied detect text from the scanned documents. In recent years, researchers have focused on deep learning approaches for digitization of handwritten data due to its better performance and adaptation capability [5].

Main objective of this work is to propose a model to detect text from hand written documents for which it is difficult to achieve high accuracy because of not a common size and shape. The focus behind this research work is to improve performance of OCR for detecting text from handwritten documents. For this purposed OpenCV approach is used and J&M model is proposed. This research work will be providing a mechanism to efficiently retrieve the information from handwritten documents which can be stored and utilize in future.

This research paper is divided into different parts such as section 2 is about literature review of exiting work and section 3 describes the adopted research. Section 4 is the experimental results followed by conclusion and future scope of this work which is described in Section 5.

## 2. Related work

In early years, researchers have explored different approaches for text detection in different complex backgrounds by advanced machine learning and optimization methods such as unsupervised machine learning [6], convolutional neural networks (CNN) [7, 8], deformable part-based models (DPMs) [9], belief propagation [10] and conditional ran-

* Corresponding author.
*E-mail address:* wss.manojkumar@gmail.com (M. Kumar).

https://doi.org/10.1016/j.gltp.2022.03.028

Available online 2 April 2022

dom fields (CRF) [11,12]. In 2014, a cost-effective model was given to recognize hand-written character and comparison was done between pixel based and feature based methods. From results, it was analyzed that pixel-based methods perform better for large datasets [13]. In 2015, a comprehensive literature survey was done for these text detection and recognition approaches. They have also summarized all problems and subproblems along with the application and challenges [14]. A single character CNN was used to propose a sliding window text detection approach across different regions [15]. After that, extra bounding box regression CNN was used for text detection [16]. Manwatkar and Singh [17] have reviewed various methods to extract characters from images and described different image processing techniques for extracting information from the scanned images. In 2017, a STN-OCR model was proposed using semi supervised neural networks for text detection. The model was investigated for different tasks and on public benchmark dataset [18]. In 2018, work was done to detect text from colored images. First text segmentation was done and after that classifier was used to recognize the images which was containing text. This methodology performs 20% better Tesseract OCR [19]. A model was proposed for automatic number plate recognition using OpenCV approach with 60% successive rate and average processing time was 0.2second to complete the entire process [20]. Otsu's segmentation approach was used segmentation and skew detection for Hough transformation in 2018 to achieve approximate 93% of accuracy [21]. In 2018, CNN method was used for image recognition on MNIST and CIFAR-10 dataset and higher degree of accuracy was achieved [22]. In 2019, work was also done to use OCR technique for extracting text from different social media images [23]. The importance and different steps of handwritten text recognition is also discussed by Karthick, et.al [24]. Later, CNN approach was applied for a proposed model which can identify text from different kind of images [25]. A survey was also done on scene text detection and recognition using different methods proposed in year from 2014-2019 and performances of those approaches were also analyzed on different datasets [26]. A deep learning approach was discussed for text detection from biomedical images which was giving higher precision of 98.6% [27]. Neural network was also applied in 2019 for scene text detection which was exploring each character and affinity between characters. For experimental work CTW-1500 dataset was used [28]. A novel text detector approach was proposed in 2020 based on weakly supervised learning. Performance of this model was close to the fully supervised model on ICDAR2015, CTW1500, Total-Text, and MSRA-TD5000 [29]. In 2020, a pretrained language representation model (BioBERT) was proposed for biomedical text mining and results were analyzed on different dataset [30]. Research was also done to recognize text from documents written in Urdu language. For which TSDNN neural network approach was proposed [31]. In 2021, OpenCV approach was applied on QR code and text images to extract the data which was used for robot navigation [32]. A new real scene text detection framework was proposed in 2021 where pretrained ResNeT-90 was employed which shows better results in compare of three different algorithms [33]. An open-source toolbox (MMOCR) was given in 2021for text detection and recognition which was also compared with other open-source OCR projects [34]. Recently in 2022, OCR and Graph Convolutional Network (GCN) approaches are also applied for paragraph detection from document images [35]. In addition to that Deep learning has been also applied for E-Commerce picture text recognition where a new approach (GTNeT) was proposed to improve the classification accuracy [36].

From this literature review, it has been analyzed that ample research has been done for text detection from scanned images and documents but still there is lot of scope for identifying the text accurately from handwritten text.

## 3. Methodology

In this section methodology of proposed J&M model is described. For storing information from handwritten documents in digital format, firstly document is scanned and then it is stored as images but searching of the specific content from these documents is a difficult task. For providing the solution of this problem, this research work focuses on training and testing the neural network model to conduct Document Image Analysis. The complete methodology of research work is shown in Fig. 1.

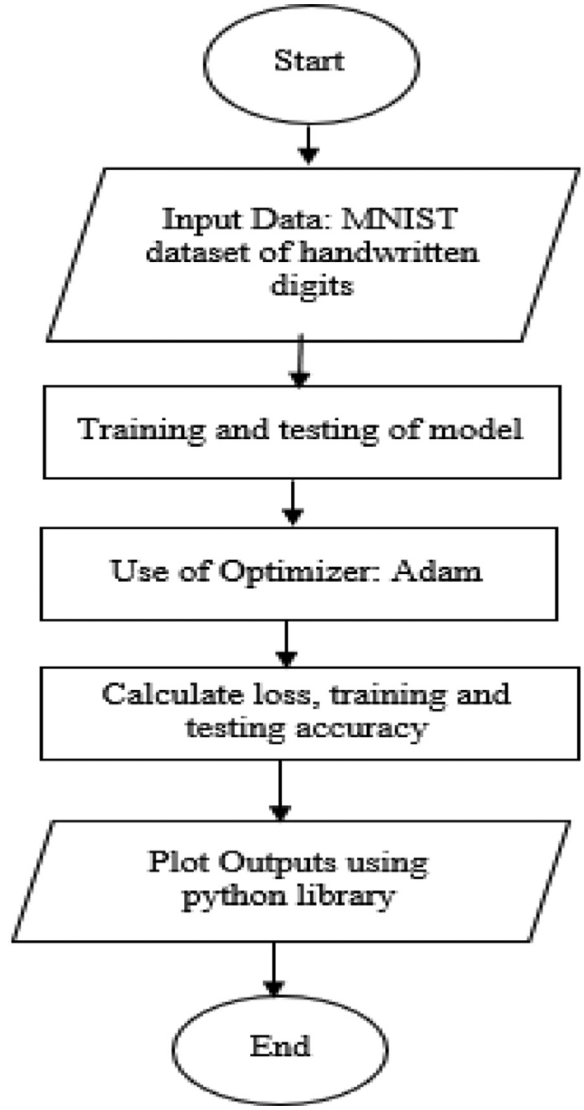


**Fig. 1.** Research methodology for proposed J&M model.

### 3.1. Input data

In this research work MNIST database of handwritten digits is used [37]. This dataset is having a training set of 60,000 images and a test set of 10,000 images. The digits used in this dataset is size-normalized and cantered in a fixed- size image.

### 3.2. Training and testing the model

a Create a classifier using convolution layers, the output will be of shape (1, 1, 10) for a single image to classify from the 10 different classes.
b Larger images would have larger output shape, but this will be used to find parts of it that were responsible for the largest activation, which also means the bounding boxes will be of (28,28) shape, the same as the number's images.

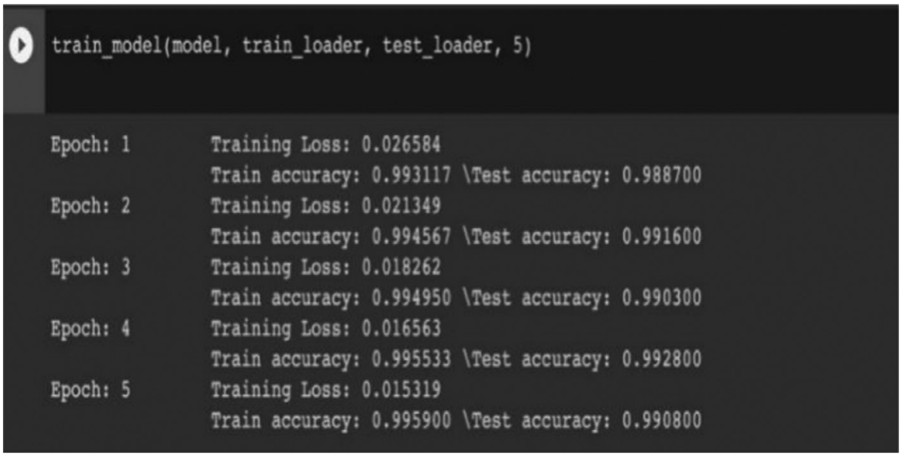


**Fig. 2.** Training loss, training and testing accuracy.

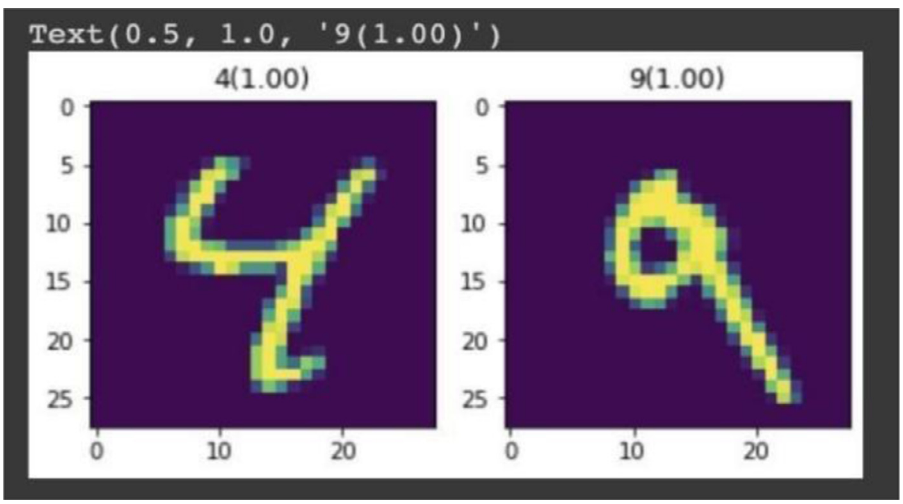


**Fig. 3.** Results obtained from randomly picked images.

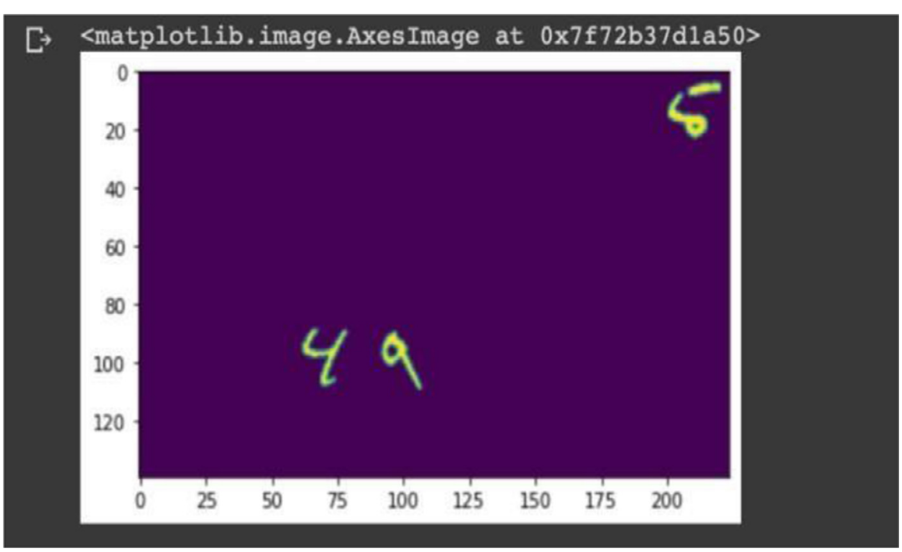


**Fig. 4.** Results obtained from randomly picked large size images.

c The classifier divides the training set into 5 subsets having approximate equal number of objects. After that his classifier is then trained 5 times and every time a single subset is excluded which will be used later on for testing.

### 3.3. Optimizer used

For optimization, Adam replacement optimization algorithm is used. It combines the best features of AdaGrad and RMSProp algorithms so that it makes possible to handle sparse gradients on noisy problems.

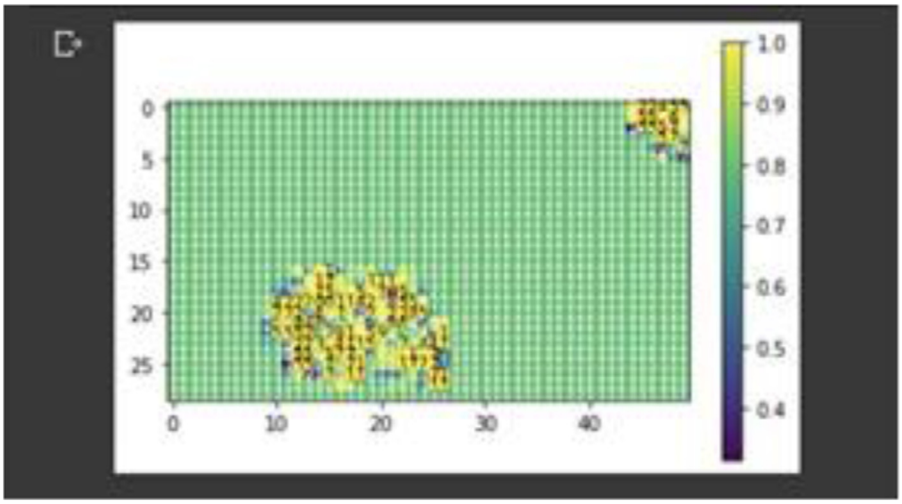


**Fig. 5.** Results obtained from randomly picked large size images.

Iterative Formulas used in Adam Optimizer is given as [38]:

$$g = \left(h_\theta(x\backsim - y\backsim)\right)x^i$$

$$m_t = \beta_1 m_{t-1} + \left(1-\beta_1\right) * g$$

$$v_t = \beta_2 v_{t-1} + \left(1-\beta_2\right) * g^2$$

$$m_t^- = \frac{m_t}{1-\beta_1^t} \quad (1)$$

$$v_t^- = \frac{v_t}{1-\beta_2^t}$$

$$\theta_j = \theta_{j-1} - m_t^- * \frac{\propto}{\sqrt{v_t^-} + \in}$$

where g is the calculated gradient, $m_t$ denotes the first moment of gradient g, $v_t$ is the second moment of gradient g, $\beta_1$ stands for the first-order moment attenuation coefficient, $\beta_2$ stands for the second moment attenuation coefficient, $\theta$ represents the parameter that needs to be solved, $m_t^-$ and $v_t^-$ describe offset correction of $m_t$ and $v_t$, respectively.

### 3.4. Calculating loss, training and testing accuracy

After finishing all steps, training, and testing accuracy are calculated along with training loss. All these results are described in Section 4 using Fig. 2.

### 3.5. Plotting the Output

Experimental results are plotted using MATLAB library. These results are shown in Section 4 in Figs. 3–5.

**Table 1**
Results obtained for training loss, training and testing accuracy.

| No. of epoch | Training loss | Training accuracy | Testing accuracy |
|---|---|---|---|
| 1 | 0.026584 | 0.993117 | 0.988700 |
| 2 | 0.021349 | 0.994567 | 0.991600 |
| 3 | 0.018262 | 0.994950 | 0.990300 |
| 4 | 0.016563 | 0.995533 | 0.992800 |
| 5 | 0.015319 | 0.995900 | 0.990800 |

## 4. Results and discussion

### 4.1. Training loss, training and testing error

Experimental results obtained from proposed J&M model for training and testing accuracy along with training loss is given in Table 1 and results are shown in Fig. 2.

### 4.2. Visualizing the results after randomly picking an image from the dataset

Results are obtained for randomly picked images from the dataset to check the accuracy of results which is shown in Fig. 3.

### 4.3. Visualizing the result in large size input images

To get more clear results, large size images are also used from the dataset to check the accuracy of results which is shown in Fig. 4.

### 4.4. Plotting the parts in the input image which activated the most into output

.

**Table 2**
Results comparison with other existing methods.

| S.No | Feature extraction method | Classifier used in methods | Accuracy (%) | Source |
|---|---|---|---|---|
| 1 | R-HOG | SVM | 95.64 | [39] |
| 2 | HOG | SVM | 81 | [40] |
| 3 | HOG | SVM | 83.60 | [41] |
| 4 | HOG | ANN | 97.33 | [42] |
| 5 | CCH | SVM | 98.48 | [43] |
| 6 | CNN | CNN+SVM | 94.40 | [44] |
| 7 | CNN | SVM, KNN, Random Forest, Multilayer Perception, CNN | 96-98 | [45] |
| 8 | Proposed(J&M) OpenCV with Adam Optimizer | | 99 | |

### 4.5. Comparison with another methods

To compare the performance of proposed model, comparison has been done with another existing research work which is described in Table 2.

## 5. Conclusion and future scope

In this research work OpenCV approach was used for proposing a model (J&M model) to detect text from handwritten documents. The resultant model that is trained and tested for Optical Character Recognition For this purpose, the proposed model was trained using MNSIT dataset of handwritten digits. For training purpose, 60,000 images and for testing 10,000 images were used. After finishing experimental work, the model provides 99.59% of training accuracy 99.0% of testing accuracy along with the training loss of 1.5% which was high in compared to existing research work. This research work can be utilized for obtaining information for handwritten documents which can be stored for future uses. This process will save the time for searching the documents also because all data is gathered at one place. In future, this research work can be extended by working on a larger dataset of handwritten documents and the proposed model can be applied in different application areas. This work can be extended to reduce the training loss by providing more training to the model so that accuracy can be improved more. Eqn 1.

## References


[1] SK. Adyanthaya, Text recognition from images: a study, Int. J. Eng. Res. Technol. (IJERT) 8 (13) (2020).

[2] M Brisinello, R Grbić, D Stefanović, R. PečkaiKovac, Optical character recognition on images with colorful background, in: proceedings of IEEE 8th International Conference on Consumer Electronics, Berlin, 2018, pp. 1–6.

[3] B.D. Parameshachari, K.M. Keerthi, T.R. Kruthika, A. Melvina, R. Pallavi, K.S. Poonam, Intelligent human free sewage alerting and monitoring system, in: 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), Atlantis Press, 2021, September, pp. 480–486.

[4] CC Tappert, CY Suen, T. Wakahara, The state of the art in online handwriting recognition, IEEE Trans. Pattern Anal. Mach. Intell. 12 (8) (1990) 787–808.

[5] T.G. Nguyen, T.V. Phan, D.T. Hoang, T.N. Nguyen, C. So-In, Efficient SDN-based traffic monitoring in IoT networks with double deep Q-network, in: International conference on computational data and social networks, Springer, Cham, 2020, pp. 26–38.

[6] MA Radwan, M Khalil, H. Abbas, Neural networks pipeline for offline machine printed Arabic OCR, Neural Processing Letters 48 (2) (2018) 2018.

[7] K. Yu, L. Lin, M. Alazab, L. Tan, B. Gu, Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system, IEEE Trans. Intell. Transp. Syst. 22 (7) (2020) 4337–4347.

[8] T Breuel, A Ul-Hasan, M Al-Azawi, F. Shafait, High-performance OCR for printed English and Fraktur using LSTM networks, 12th international conference on document analysis and recognition, 2013.

[9] R.K. Dash, T.N. Nguyen, K. Cengiz, A. Sharma, Fine-tuned support vector regression model for stock predictions, Neural Comput. Appl. (2021) 1–15.

[10] A Coates, B Carpenter, C Case, S Satheesh, B Suresh, TT Wang, AY. Ng, Text detection and character recognition in scene images with unsupervised feature learning, in: 2011 International Conference on Document Analysis, 2011, pp. 440–445.

[11] Z. Guo, K. Yu, A. Jolfaei, A.K. Bashir, A.O. Almagrabi, N. Kumar, Fuzzy Detection System for Rumors Through Explainable Adaptive Learning, IEEE Trans. Fuzzy Syst. 29 (12) (2021) 3650–3664.

[12] D Wu Wang, A Coates, A Ng, End-to-end text recognition with convolutional neural networks, in: Proceedings of the 21st international conference on pattern recognition (ICPR2012), IEEE, 2012, pp. 3304–3308.

[13] Krishnaiah, A., & Divakarachari, P. B. Automatic music mood classification using multi-class support vector machine based on hybrid spectral features.

[14] Y Zhu, J Sun, S. Naoi, Recognizing natural scene characters by convolutional neural network and bimodal image enhancement, in: International Workshop on Camera-Based Document Analysis and Recognition, Springer, Berlin, 2011, pp. 69–82.

[15] C Shi, C Wang, B Xiao, Y Zhang, S Gao, ZZ. Zhang, Scene text recognition using part-based tree-structured character detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2961–2968.

[16] D.L. Vu, T.K. Nguyen, T.V. Nguyen, T.N. Nguyen, F. Massacci, P.H. Phung, A convolutional transformation network for malware classification, in: 2019 6th NAFOSTED conference on information and computer science (NICS), IEEE, 2019, December, pp. 234–239.

[17] JJ Weinman, EL Miller, AR. Hanson, Scene text recognition using similarity and a lexicon with sparse belief propagation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (10) (2009) 1733–1746 2009.

[18] Z. Guo, K. Yu, Y. Li, G. Srivastava, J.C.W. Lin, Deep learning-embedded social internet of things for ambiguity-aware social recommendations, IEEE Trans. Netw. Sci. Eng. (2021).

[19] YF Pan, X Hou, C. Liu, Text localization in natural scene images based on conditional random field, in: 2009 10th International Conference on Document Analysis and Recognition, IEEE, 2009, pp. 6–10.

[20] H Zhang, C Yang, et al., An improved scene text extraction method using conditional random field and optical character recognition, in: 2011 International Conference on Document Analysis and Recognition, IEEE, 2011, pp. 708–712.

[21] N Chauhan, M Sharma, P. Singh, A hybrid approach towards cost effective model for handwritten character recognition, Int. J. Comput. Appl. 95 (14) (2014) 36–39.

[22] Y Qixiang, D. Doermann, Text detection and recognition in imagery: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 37 (7) (2014) 1480–1500.

[23] M Jaderberg, A Vedaldi, A. Zissermerg, Deep features for text spotting, Computer Vision - ECCV 2014, number 8692 Lect. Notes Comput. Sci. (2014) 512–528.

[24] M Jaderberg, K Simonyan, A Vedaldi, A. Zissermerg, Reading text in the wild with convolutional neural networks, Int. J. Comput. Vision 116 (1) (2014) 1–20.

[25] PM Manwatkar, KR. Singh, A technical review on text recognition from images, in: IEEE 9th International Conference on Intelligent Systems and Control, 2015, pp. 1–5.

[26] C Bartz, H Yang, C. Meinel, STN-OCR:A single neural network for text detection and text recognition, arXiv preprint (2017) arXiv:1707.08831.

[27] M Brisinello, R Grbic, D Stefan, R. PeckaiKovac, Optical character recognition on images with colorful background, in: 8th International Conference on Consumer Electronics, IEEE, Berlin, 2018, pp. 1–6.

[28] AS Agbemenu, J Yankey, EO. Addo, An automatic number plate recognition system using opencv and tesseract ocr engine, Int. J. Comput. Appl. 180 (43) (2018) 1–5.

[29] N Agarwal, A. Kaur, An algorithm approach for text recognition from printed/typed text images, in: 8th International Conference on Cloud Computing, Data Science & Engineering, 2018, pp. 876–879.

[30] R Chauhan, KK Ghanshala, RC. Joshi, Convolutional neural network (CNN) for image detection and recognition, in: First International Conference on Secure Cyber Computing and Communication (ICSCCC), IEEE, 2018, pp. 278–282.

[31] MS Akopyan, O Belyaeva, TP Plechov, DY. Turdakov, Text recognition on images from social media, in: 2019 Ivannikov Memorial Workshop (IVMEM), IEEE, 2019, pp. 3–6.

[32] K Karthick, K Ravindrakumar, R Francis, S. Ilankannan, Steps involved in text recognition and recent research in OCR; a study, Int. J. Recent Technol. Eng. 8 (1) (2019) 277–3878.

[33] A Shrivastava, J Amudha, D Gupta, K. Sharma, Deep learning model for text recognition in images, in: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2019, pp. 1–6.

[34] Xiyan Liu, G Meng, C. Pan, Scene text detection and recognition with advances in deep learning: a survey, Int. J. Doc. Anal. Recognit. (IJDAR) 22 (2) (2019) 143–162.

[35] Wenyuan Xue, L Qingyong, X. Qiyuan, Text detection and recognition for images of medical laboratory reports with a deep learning approach, IEEE Access 8 (2019) 407–416.

[36] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, Hwalsuk Lee, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9365–9374.

[37] Computational intelligence in data-driven modelling and its engineering applications 2020.