9-2022

Deep learning-based text recognition of agricultural regulatory document

Hua Leong FWA
Singapore Management University, hlfwa@smu.edu.sg

Farn Haur CHAN
Syngenta Asia Pacific Pte Ltd

Citation

FWA, Hua Leong and CHAN, Farn Haur. Deep learning-based text recognition of agricultural regulatory document. (2022). *Advances in Computational Collective Intelligence: 14th International Conference, ICCI 2022, Hammamet, Tunisia, September 28-30: Proceedings*. 1653, 223-234.
Available at: https://ink.library.smu.edu.sg/sis_research/7334

Deep learning-based text recognition of agricultural regulatory document

First Author[1][0000−1111−2222−3333], Second Author[2,3][1111−2222−3333−4444], and Third Author[3][2222−−3333−4444−5555]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[3] ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

Abstract. In this study, an OCR system based on deep learning techniques was deployed to digitize scanned agricultural regulatory documents comprising of certificates and labels. Recognition of the certificates and labels is challenging as they are scanned images of the hard copy form and the layout and size of the text as well as the languages vary between the various countries (due to diverse regulatory requirements). We evaluated and compared between various state-of-the-art deep learning-based text detection and recognition model as well as a packaged OCR library – Tesseract. We then adopted a two-stage approach comprising of text detection using Character Region Awareness For Text (CRAFT) followed by recognition using OCR branch of a multi-lingual text recognition algorithm E2E-MLT. A sliding windows text matcher is used to enhance the extraction of the required information such as trade names, active ingredients and crops. Initial evaluation revealed that the system performs well with a high accuracy of 91.9% for the recognition of trade names in certificates and labels and the system is currently deployed for use in a client site in Philippines.

Keywords: deep learning · text detection · optical character recognition · regulatory document.

1 Introduction

In many industries, the conversion of company assets from non-machine readable into machine readable form is a prerequisite of their journey towards digitalization – the use of digital technologies to transform their business model and provide new opportunities for optimizing their business operations and/or discovering new revenue streams. Optical Character Recognition (OCR) is frequently employed to convert scanned documents in images and pdf formats into machine-readable form for further data processing.

Our client is a leading provider of agricultural science and technology who specializes in the development and provision of seeds and crop protection products to farmers. Rampant use of pesticides as a convenient way of eradicating pests found on crops and reports of possible impact of human health in the early years have prompted the government of various crop producing nations to regulate and control the use of seed and crop protection products such as pesticides and herbicides [6]. These regulatory measures include the submission of an application to register for pesticide or herbicide use accompanied by the relevant test data. A draft of the product label which lists the composition of the product, application and use of the product must also be filed with the regulatory authority.

Upon the approval of the registration, the regulatory authority will then issue a signed certificate which validates the product to be ready for sale and use in the relevant markets for a specified duration. These labels and certificates contain important information such as the active ingredient, the crops and pests which the product is targeted to be use on and the mode and rate of application of the product e.t.c. The labels and certificates can be in soft copy but non-textual format e.g. graphical formats or can be scanned images of the actual hard copy, thus necessitating the use of OCR techniques to extract the textual information within the certificates and labels. In addition, the layout and size of the text and language within the certificates and labels varies between the different countries due to diverse languages and regulatory requirements. This issue of differing regulatory standards and languages is especially pervasive among the countries within Asia where this research is targeted for.

The early optical character recognition efforts involve the use of hand-crafted low-level features e.g. stroke width, gradient, texture e.t.c. [4, 14, 7]. In most contexts, tedious image pre-processing and post-processing are also required to improve the accuracy of text detection and recognition. This reduces the robustness in the use of such techniques for diverse images e.g. images with diverse fonts, layouts and orientations, multilingual texts and complex backgrounds. These attributes are prevalent in scene text or text that occurs in our natural environment. As opposed to text recognition of type-written or well-formatted documents with high document image quality, achieving high recognition accuracy for scene text is a challenging endeavor as it is characterized by more variability and noise [28].

In recent years, deep learning techniques hold great promise for both image as well as text detection and recognition. The advent of deep learning techniques has contributed to state-of-the-art results on challenging computer vision problems such as object detection and classification[15, 22, 9, 25]. The deep learning models for object detection and classification are similarly adapted for use in text detection and recognition e.g. the use of Convolutional Neural Network model LeNet [16] for handwritten digit recognition. Several deep learning architectures involving the use of Convolutional Neural Networks (CNNs) have evolved along the years with deeper layers and enhanced accuracy for scene text detection and recognition − localizing and recognizing text that occurs in natural images. In contrast to the use of traditional image processing techniques, deep learning techniques eliminate the need for laborious creation and testing of low-level hand-crafted features which thus lead to better model generalizability.

We seek to evaluate and compare between various state-of-the-art deep learning based text detection and recogition models and a pre-packaged OCR library - Tesseract for an automated extraction of essential information from agricultural regulatory documents in this study. The main contributions of this work are as follows:

- We evaluated state-of-the-art deep learning models for text detection and recognition for use in a real-life industrial application.
- We detailed the techniques that we used for enhancing the accuracy of capturing the essential information from regulatory documents i.e. certificates and labels.
- Lastly, we presented a real-life implementation of regulatory information capturing system with a high level of accuracy.

2 Related Studies

2.1 Text Detection and Recognition

OCR comprises of first distinguishing the text from non-text region within an image (text region detection) and then recognizing the text content (text recognition). Some studies on text detection and recognition adopted a two-stage approach where text region detection is first performed for prediction of the bounding boxes before passing the image cropped using the predicted bounding box to a separate text recognition model. There were also studies which adopted a single stage approach where both text detection and recognition were trained in a single end-to-end pass. A single stage approach models a joint loss function for both localization and recognition and this is postulated to result in a higher accuracy of recognizing the text. More studies are however required to justify this [12]. In our study, we selected a dual stage model for our implementation.

Text detection using deep learning techniques can be divided into Regression-based, Segmentation/link-based and Character-level detectors. In regression-based methods, horizontal/quadrangular boxes are first placed over text instances [20, 19]. Iterative regression of the bounding boxes is then applied to generate coordinates of bounding boxes that enclosed the text. A structural limitation of regression-based methods is with their inability to capture all possible text bounding shapes that occur in the wild. Segmentation/link-based methods overcome this limitation by breaking text instances into multiple parts and then linking these parts using some features to detect the final text [26]. The Efficient And Accurate Scene Text detection (EAST) [27] is a segmentation based method that generates multi-oriented text predictions by feeding image pixels directly into a Fully Convolutional Neural Network (FCN) architecture. The model then outputs multiple channels of score map (representing the confidence of the predicted geometry shape) and the geometry shape coordinates. Character level detectors detect text with characters as its unit instead of words. The Character Region Awareness For Text (CRAFT) [1] is one such model which used a FCN to output character region score and affinity score for localizing individual character and linking the detected characters to text respectively. From the published results, CRAFT demonstrated robustness in detecting text with varying sizes including curved or deformed text.

Methods for text recognition usually consist of a feature extractor that maps the text instance image into a representation with attributes that are relevant for text recognition. Convolutional Neural Networks (CNNs) are widely used in this feature representation stage [5]. After extraction of the text features, the prediction stage then estimates the target string sequence from the extracted features. The two major techniques that are applied here are the Connectionist Temporal Classification (CTC) [8] and the Attention mechanism [2]. CTC converts the features made by CNNs into a target string sequence by calculating the conditional probability. In CTC, the maximum likelihood of the output sequence is computed by summing over the probabilities of all possible input-output sequence alignments. With the insertion of a blank symbol to denote repeated occurrence of a character in CTC, there is no need for prior alignment between the input and target output sequences. Attention on the other hand, learns the alignment between the input and target output sequences from the history of the target characters and the encoded features. Both CTC and Attention are widely used in scene text recognition deep learning architectures.

Industrial Applications of OCR The study by Kim et al. [13] proposes a method of symbols and text in Piping and Instrumentation Diagrams (P&ID) within the process industry. The need for digitizing P&IDs arises as many P&IDs especially for older plants exist in image form. The authors first constructed a P&ID dataset consisting of 386 symbols relating to piping and instrumentation before passing the dataset to a deep learning model for symbol recognition. For the text portion, the authors used easyOCR for text region detection and Tesseract [21] for text recognition. They justified that this combination provided better performance. The proposed system achieved a precision of 0.9718 and recall of 0.9827 for symbol recognition. The precision and recall for text recognition are 0.9386 and 0.9175. The authors acknowledged that text recognition in P&IDs is less challenging than that in natural images as P&IDs are properly formatted document with no complex background.

A system for the detection and recognition of embossed codes printed on tyre sidewalls for automated tyre condition monitoring was proposed by Kazmi et al. [11]. The authors first applied circular Hough transformation to unwarp text arches into rectangular shaped text. A two-stage approach was adopted where a histogram of gradient features formulated from the tyre images, was passed into a CNN cascade for localizing the text codes before passing the cropped image into another cascade of CNNs for text recognition. In this study, an average accuracy of 86% was attained with the proposed system as opposed to close to zero recognition when using Tesseract. The authors thus postulated that the use of packaged OCR library may not be adequate for all OCR contexts.

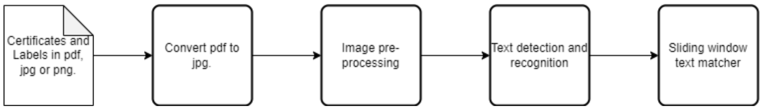

Certificates and Labels in pdf, jpg or png.
Convert pdf to jpg.
Image pre-processing
Text detection and recognition
Sliding window text matcher

Fig. 1. Overall processing workflow for recognizing and extract essential text information

3 Methodology

Fig 1 shows the overall processing workflow for recognizing and extracting the essential information from product certificates and labels.

3.1 Image pre-processing

We pre-processed the documents by first converting them from pdf into image files. To achieve higher accuracy of detection, we converted the images to greyscale and applied adaptive thresholding. The quality of the document scans varies and some of these scanned document may be slanted or oriented in the wrong direction e.g. rotated 90 degrees either clockwise or anti-clockwise. Automatic correction of the document orientation is necessary to ensure that the subsequent text detection and recognition is as accurate as possible. We applied Canny edge detection and Hough transform [23] to calculate the correction angle for proper orientation of the document. Finally, we also resized the document image to constraint it to a maximum of 1280 pixels in either the horizontal (for width larger than height) or vertical dimension (for height larger than width) while maintaining the original aspect ratio.

3.2 Evaluation of text detection and recognition models

We employed CRAFT as the text detection method for this study. From our initial evaluation comparing CRAFT to EAST, we found that CRAFT performed better with more text detected and lesser overlaps in the predicted bounding boxes. We used a pre-trained CRAFT model for the text detection phase in this project.

For text recognition, we experimented with 3 methods. For the first method, we used the text recognition branch of E2E-MLT [3] which is made up of deep convolutional layers. For the second method, we used Show, Attend and Read model (SAR) [18] which makes use of ResNet [9] and attention-based Long Short Term Memory (LSTM) [10] encoder-decoder architecture. For the last model, we used Tesseract, an open-source text recognition library from Google [24]. For E2E-MLT and SAR models, we used the pre-trained weights provided by their respective authors. In E2E-MLT, the original model is an end-to-end single stage model which optimized both text localization and text recognition with a joint loss function. The original E2E-MLT generates and filters text proposals for detecting and locating text within an image. We replaced the text localization branch of E2E-MLT with CRAFT and used only the text recognition branch of the model. Our rationale is since it has not been proven conclusively that an end-to-end text detection and recognition model works better than a dual stage one, this dual stage design allows us the flexibility of evaluating and selecting the best of breed models separately for the text detection and recognition phases.

Table 1. Character error rates (CER) of models

Model	CER
E2E-MLT	0.12
SAR	0.32
Tesseract	0.10

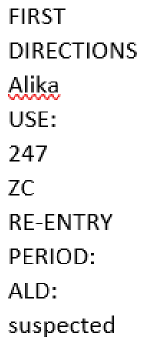
To evaluate the performance of the 3 text recognition models, we randomly selected 5 certificates and 5 labels from our dataset. For E2E-MLT and SAR, we passed the certificates and label images through CRAFT model for prediction of the bounding boxes which is then used for cropping out the detected text. The cropped text images were further passed to both E2E-MLT and SAR for text recognition. For Tesseract, we passed the certificates and label images directly into the library for both text detection and recognition. We compared the character error rate for the first 30 detected words with their ground truth for each label and certificate. The character error rate is a metric that is used to evaluate OCR output quality. The results are shown in Table 1.

As can be seen in Table 1, the CER for Tesseract is the best at 10% while the CER for E2E-MLT is comparable to that of Tesseract at 12%. SAR's CER is rather high at 32%. In addition, we also calculated the percentage of instances of non-text regions detected as a percentage of the total detection instances for both CRAFT and Tesseract. We noted that Tesseract detected a higher number of instances of non-text regions at 26.8% while that for CRAFT is just 6.9%. A higher incidence of false positives in the detection of text region would result in higher inaccuracies in subsequent text recognition on top of the cost of additional processing time. Thus, considering both model CER and false positives of text detection, we selected E2E-MLT (with CRAFT for text detection) as the final model for our implementation.

4 Models

4.1 Character Region Awareness For Text (CRAFT)

CRAFT is a fully convolutional network architecture which uses VGG-16 as its backbone. VGG-16 is a feature extraction architecture that encodes an image input into a certain feature representation. The decoding segment of CRAFT also uses skip connections to aggregate lower-level features. CRAFT predicts two



FIRST
DIRECTIONS
Alika
USE:
247
ZC
RE-ENTRY
PERIOD:
ALD:
suspected

Fig. 2. Sample OCR text output

scores for each character – region score which localizes the character and affinity score, a measure of association of one character with another. The affinity score is used for combining the characters together into words. The combination of these two scores then defines the bounding boxes that enclose the detected text.

4.2 E2E-MLT OCR branch

The OCR branch of the E2E-MLT is a FCN module for multi-language text recognition. The bounding boxes predicted by CRAFT are used to crop out the word level image segments and the cropped images are then passed into the FCN module for recognition of the specific text. The cropped image passed into the FCN module are scaled to a $\bar{W}$ x 40 x C where height (h) of the cropped image is fixed at 40 pixels and $\bar{W}$=w$\frac{H'}{h}$. The FCN module then outputs a matrix of size $\frac{\bar{W}}{4}$ x $\left|\hat{A}\right|$, where $\hat{A}$ is the union of characters in all languages. The loss for text recognition is computed using CTC. E2E-MLT OCR supports 8 different languages – Arabic, Bangla, Chinese, Japanese, Korean and Latin.

Algorithm 1 Sliding window text matcher algorithm

```
procedure WindowedTextMatcher($wordlist, cutoffscore$)
    $w \leftarrow 1$                                          ▷ w is the window width
    while not end of list do
        while $w <= 5$ do                                 ▷ Loop for window size 1 to 5
            Combine k words to form new word separated by space
            while not end of global dictionary do
                Calculate Levenshtein ratio (lr) between new word(s) and the word(s) in global dictionary
                if $lr \geq cutoffscore$ then
                    return $word$
    return $null$
```

4.3 Sliding windows text matcher

The outputs from CRAFT and E2E-MLT models comprises of a list of words detected from the individual certificate or label. From the list of words, it is still not possible to identify the trade name, active ingredient, crops and pests. One technique to extract the required information e.g. trade name is to annotate a few possible locations where the trade name is found on the certificates or labels. However, this limits both the generalizability and reliability of the system as a change in the document format or layout would result in non-extraction. Most of the useful text that we need to extract for feeding into our client's knowledge management system relate to the trade name, active ingredient and crops and pests. Some trade names, crops and pests also consist of multiple words joined together e.g. Alika 247 ZC. In addition, the terms or words that make up the trade name may not be in consecutive locations due to diverse font sizes and orientations. A sample of a part of the output from the text recognition model is shown in Fig 2.

From the sample OCR text output, it would not be possible to pick out the trade name "Alika 247 ZC" as it is split into 3 different words and at non-consecutive locations. To resolve this, we first extracted a global dictionary of trade names, active ingredients, crops and pests from both the company's internal sources as well as from the company's subscriptions of external databases. A sliding windows text matcher is then applied to match the global dictionary against the detected text.

The formula to calculate Levenshtein Ratio (LR) is given below.

$$LR = (len(S) + len(D) - LD)/(len(S) + len(D)) \tag{1}$$

Where LD is the Levenshtein Distance. LR is used as it normalizes for different word length, giving a score between 0 and 100 where 100 denotes a perfect match between the source and destination word. Levenshtein Distance(LD) [17] is given by

$$lev_{a,b}(i,j) = \begin{cases} max(i,j) & \text{if min(i,j)=0} \\ min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1 \end{cases} & \text{otherwise} \end{cases} \tag{2}$$

Table 2. Recognition accuracy of trade name, active ingredients and crops

Document	Trade Name	Active Ingredients	Crops
Certs	94.90%	82.00%	75.80%
Labels	88.90%	64.30%	95.20%
Overall	91.90%	73.15%	85.50%

5 Results

With the proposed sliding windows text matcher, we refined the extraction of trade names, active ingredients and crops from a total of 39 certificates and 9 labels. These certificates and labels are provided by our client in Philippines (where our proposed solution has been deployed).

As seen in Table 2, the recognition of trade name from certificates is high at 94.9% accuracy. For labels, the accuracy for recognition of trade name is lower at 88.9% accuracy. On closer examination, the diverse font styles and text sizes for the text found in product labels made it more challenging to accurately detect the text bounding boxes and thus lowered the accuracy of text recognition. Aggregating across certificates and label, the recognition accuracy for trade name is the highest at 91.9% followed by that for crops at 85.5% and lastly for active ingredients at 73.15%. We did not measure the recognition accuracy of pests as the global database currently only contains the pests' scientific names and not their more commonly known English names. However, the pest names that are found on labels and certificates are their common English names and not the scientific names. Our initial investigation also revealed that there were overlaps between the English and scientific pest names such that a pest common name might be associated with multiple scientific names. To date, we are still sourcing for a reliable database that maps the pests' scientific names to their common English names.

6 Real world deployment

We have deployed the proposed system in one of our client's site in Philippines and used it for the automated recognition of agricultural regulatory documents - certificates and product labels. Although the current volume is not huge (around 10 to 20 certificates and labels per month), the recognition accuracy is close to what is detailed in Table 2. Furthermore, our client has plans to scale this system to be deployed to multiple sites within Asia. With the deployment of this system, the user will not need to key in detail such as trade names, active ingredients and crops into a separate knowledge management system manually. This would not only conserve man effort but also reduce input errors. We estimate that it takes about half an hour for a single user to key in the details listed in a single certificate or label. For an average volume of 100 certificates or labels per month, our system can achieve 50 man hours savings and this will be more significant when the use of the system is further scaled up.

7 Conclusion

In this study, we have detailed the design of an OCR system based on deep learning techniques that is used for the recognition of agricultural regulatory document. We have also evaluated the use of state-of-the-art deep learning based text detection and recognition algorithms as well as a packaged OCR library, Tesseract for the recognition of the regulatory certificates and labels. With the use of CRAFT for text detection and E2E-MLT OCR for text recognition, we managed to achieve an accuracy of 91.9%, 73.15% and 85.5% for the recognition of trade name, active ingredients and crops. To avert the need to annotate possible locations where the required term is located within the regulatory document and to achieve a high generalizability and reliability of extraction, we used a sliding window text matcher to match against a global dictionary of the required terms. The proposed system is currently deployed for use in a client site in Philippines and the user has plans to scale up the use of the system.

In order to scale the use of the system in other sites within Asia, an important future task is the detection and recognition of multi-lingual text. E2E-MLT can currently support 8 languages but we intend to explore other multi-lingual OCR techniques and apply fine-tuning to enhance the recognition accuracy.

References

1. Baek, Y., Lee, B., Han, D., Yun, S., Lee, H.: Character Region Awareness for Text Detection. arXiv:1904.01941 [cs] (Apr 2019), http://arxiv.org/abs/1904.01941, arXiv: 1904.01941
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
3. Bušta, M., Patel, Y., Matas, J.: E2e-mlt-an unconstrained end-to-end method for multi-language scene text. In: Asian Conference on Computer Vision. pp. 127–143. Springer (2018)
4. Chen, X., Yuille, A.L.: Detecting and reading text in natural scenes. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 2, pp. II–II. IEEE (2004)
5. Chen, X., Jin, L., Zhu, Y., Luo, C., Wang, T.: Text Recognition in the Wild: A Survey. arXiv:2005.03492 [cs] (Dec 2020), http://arxiv.org/abs/2005.03492, arXiv: 2005.03492
6. Gaston, C.P.: Pesticide Regulatory Policies of Selected Countries in Asia p. 35
7. Gllavata, J., Ewerth, R., Freisleben, B.: Text detection in images based on unsupervised classification of high-frequency wavelet coefficients. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. vol. 1, pp. 425–428. IEEE (2004)
8. Graves, A., Fernandez, S., Gomez, F., Schmidhuber, J.: Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks p. 8
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
11. Kazmi, W., Nabney, I., Vogiatzis, G., Rose, P., Codd, A.: An Efficient Industrial System for Vehicle Tyre (Tire) Detection and Text Recognition Using Deep Learning. IEEE Transactions on Intelligent Transportation Systems **22**(2), 1264–1275 (Feb 2021). https://doi.org/10.1109/TITS.2020.2967316, https://ieeexplore.ieee.org/document/8968735/

12. Khan, T., Sarkar, R., Mollah, A.F.: Deep learning approaches to scene text detection: a comprehensive review. Artificial Intelligence Review **54**(5), 3239–3298 (Jun 2021). https://doi.org/10.1007/s10462-020-09930-6, https://link.springer.com/10.1007/s10462-020-09930-6
13. Kim, H., Lee, W., Kim, M., Moon, Y., Lee, T., Cho, M., Mun, D.: 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 (Nov 2021). https://doi.org/10.1016/j.eswa.2021.115337, https://linkinghub.elsevier.com/retrieve/pii/S0957417421007661
14. Kim, K.I., Jung, K., Kim, J.H.: Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(12), 1631–1639 (2003)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)
16. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (Nov 1998). https://doi.org/10.1109/5.726791, http://ieeexplore.ieee.org/document/726791/
17. Levenshtein, V.I., et al.: Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet physics doklady. vol. 10, pp. 707–710. Soviet Union (1966)
18. Li, H., Wang, P., Shen, C., Zhang, G.: Show, Attend and Read: A Simple and Strong Baseline for Irregular Text Recognition. arXiv:1811.00751 [cs] (Mar 2019), http://arxiv.org/abs/1811.00751, arXiv: 1811.00751
19. Liao, M., Shi, B., Bai, X.: Textboxes++: A single-shot oriented scene text detector. IEEE transactions on image processing **27**(8), 3676–3690 (2018)
20. Liao, M., Shi, B., Bai, X., Wang, X., Liu, W.: Textboxes: A fast text detector with a single deep neural network. In: Thirty-first AAAI conference on artificial intelligence (2017)
21. Patel, C., Patel, A., Patel, D.: Optical character recognition by open source ocr tool tesseract: A case study. International Journal of Computer Applications **55**(10), 50–56 (2012)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Singh, C., Bhatia, N., Kaur, A.: Hough transform based fast skew detection and accurate skew correction methods. Pattern Recognition **41**(12), 3528–3546 (2008)
24. Smith, R.: An overview of the tesseract ocr engine. In: Ninth international conference on document analysis and recognition (ICDAR 2007). vol. 2, pp. 629–633. IEEE (2007)
25. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
26. Tian, S., Bhattacharya, U., Lu, S., Su, B., Wang, Q., Wei, X., Lu, Y., Tan, C.L.: Multilingual scene character recognition with co-occurrence of histogram of oriented gradients. Pattern Recognition **51**, 125–134 (Mar 2016). https://doi.org/10.1016/j.patcog.2015.07.009, https://linkinghub.elsevier.com/retrieve/pii/S0031320315002691
27. Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J.: East: an efficient and accurate scene text detector. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 5551–5560 (2017)

28. Zhu, Y., Yao, C., Bai, X.: Scene text detection and recognition: recent advances and future trends. Frontiers of Computer Science **10**(1), 19–36 (Feb 2016). https://doi.org/10.1007/s11704-015-4488-0, http://link.springer.com/10.1007/s11704-015-4488-0