International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-01, APR 2021

A Creative Method for Detecting and Interpreting Invoice Document

[1]Yash Roshan Pimple, [2]Bhavik Alpesh Saglani, [3]Henil Nandu, [4]Prof. Akshay Agarwal

[1,2,3]BE Student, 4Asst. Professor, Department of Information Technology, Universal College of Engineering, Vasai, India, [1]yash.pimple98@gmail.com, [2]bhaviksaglani1998@gmail.com, [3]henilnandu98@gmail.com

Abstract: Text recognition has long been a subject of computer vision science. Previous academic approaches to scene text recognition have shown positive results, but these technologies typically fail to detect text that is present in a document format. Where the text in the picture does not follow a linear pattern, traditional OCR engines usually crash .The proposed framework is intended to provide a fast and reliable text detection system for images with systematic an d spatial data that is difficult to process using conventionnel OCR engines. Our methodical improvements result in effe ctive text detection through a system that is depéndale, fast, and easy to use.

It fits best for text that is segmented in the picture and is non-linear in nature. Our implementation easily outperforms state-of-the-art text detection strategies as opposed to other text detection algorithms such as EAST text detection. In terms of financial data management, our proposed solution will be extremely useful to companies and organisations.

Keywords— text detection, invoice text detection, document OCR, openCV, image processing, deep learning

I. INTRODUCTION

An invoice is created for some kind of financial transaction, and it includes the transaction's information in a very conci se and methodical manner. Corporations are normally requi red to keep track of their transactions in an internal ledger, a nd manual labour is used to do so. This results in a waste of time and money for the company. The organization's finances are jeopardized for no apparent gain. Modern OCR engines ignore the fact that invoice text does not follow a linear structure and is spatially separated depending on related content.[1] Moreover, not all invoices have the same format. pre-defining a set of criteria for detecting the same structure That format is useless.

Text extraction from an image is traditionally done in two st eps: text identification and text recognition.

Through preprocessing the document and detecting text fields that are designed to represent a common feature within the invoice or document, the proposed framework bridges the gap between OCR engines and invoices. These components will then be interpreted independently using the Tesseract OCR [8] engine to identify the text contained in them. This whole procedure is carried out through a single pipeline, ensuring that the operation is as effective as possible.

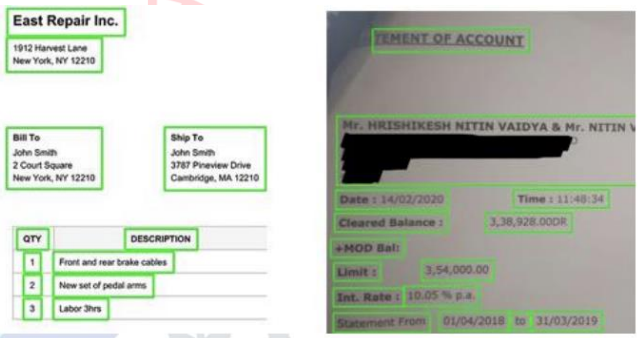

Figure 1: The proposed system can accurately output bounding boxes covering the area of text

II.RELATED WORK

Deep Learning algorithms have improved the performance of text recognition, which is essentially sliding window techniques, in recent years. [9] They all have their own improvements, which often include using very deep CNN to take advantage of high-level deep functionality and exchanging convolutional mechanisms[10], which are often used to reduce computational costs. As a result, several FCN-based approaches have been suggested, with promising findings in text detection tasks. Previously, only one of the two functions, text identification and text recognition, had been addressed. Following that, these activities are planned in order to provide full optical character recognition. Almost all of the preceding methods rely on detecting and r ecognizing scenario text. [1][2] [3] [5] .Text detection meth

 DOI : 10.35291/2454-9150.2021.0129

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-01, APR 2021

ods in recent years have largely been based on general objet detectors with various text-specific modifications. (5)

III. LITERATURE SURVEY

A Connectionist Text Proposal Network (CTPN) was suggested by Tian et al. [3] to investigate the meaning of text and to detect a text illustration in a series of fine-scale text applications.

Similarly, Shi et al. created a linking-segment approach that can distinguish multi-oriented text when retrieving a text instance in a sequence. EAST Text Detection [2] was also presented and investigated. IOU loss was used to identify several letters that resembled a word within the same parameter, and it provided the best results. To extend SSD object detector to text detection, a single-shot text detector (SSTD) [11] has recently been proposed. SSTD boosts text information by encoding geographic exposure to text into convolutional functions. Recent work on text recognition, which was inspired by speech recognition, has widely cast it as a sequence-to-sequence recognition challenge, with recurrent neural networks (RNNs) being used. He and his colleagues. [10] used Convolutional Neural Networks (CNNs) to encode a raw input image into a series of deep features, and then an RNN was added to the sequential encoding and trust mapping features, followed by connectionist temporal classification CTC. Shi and colleagues. [12] improved these system variations by making them trainable from beginning to end, resulting in substantial efficiency gains.

The scheme has recently been expanded with the addition of various attention mechanisms that can directly or indirectly encode additional character information.

IV. PROPOSED METHODOLOGY

The proposed method works for both synthesised and manually captured invoice files. The machine then processes these images by first detecting text-covered regions, creating bounding boxes around them, and then moving each bounding box through the OCR engine individually, resulting in segmented text recognition based on text categories. Modern Convolutional Neural Networks are costly to train because they require cutting-edge hardware for fast computation. Modern GPUs, such as those from Nvidia and AMD, which are marketed specifically for deep learning, can compute CNNs optimally, but at a higher cost. Furthermore, designing a CNN solely for the use of OCR has already been extensively studied, and there are currently available solutions that are basically state-of-the-art OCR programmes. Creating our own CNN for OCR purposes would be like reinventing the wheel in this situation. Instead, the system has chosen to use Tesseract, a Google open source OCR implementation. In order to preprocess the image and detect contours that contain a region of segmented text, the device also uses the OpenCV image processing library. OpenCV has a number of built-in features and applications based on recent science articles, but it's more useful for managing the system's workflow. The proposed method takes a simple approach, relying on conventional image processing methods to detect regions in an image that contain text. Several morphological operations, as well as other image recognition methods, are used in these processing techniques.

A. Image Processing

Image processing is made up of several components in which we perform several operations on the image in order to achieve a contour with the least amount of error. Morphological image processing is a collection of non-linear operations that deal with the shape or morphology of image types. Morphological operations are best suited for manipulating binary images because they depend entirely on the relative ordering of pixel values rather than their numerical values.

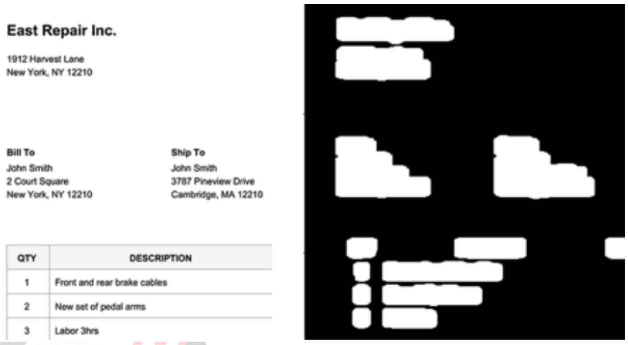

Figure 2: The binary grayscale output obtained after preprocessing the raw image

Grey-scale images can also be subjected to morphological operations such that their light transfer mechanisms are undefined and their absolute pixel values are of little or minor significance.

The suggested method dilates the picture using morph operations to enlarge the text areas and improve the finalising contours. Taking this into account, the anchor mechanism increases the ability to anticipate text components at various stages in the text detection task. A text line can be thought of as a collection of fine-scale text proposals that can be treated as a function for object detection to some extent. We expect it to operate efficiently on text recognition with various sizes and different aspect ratios by detecting a sequence of text proposals from a text line so it can include a portion of the text line and provide all of the text features.

B. Text Detection

The product of the preprocessing procedure is used to create contours around the text's key sections. These contours are created by strategically placing bounding boxes along the picture's outlined regions in the preprocessed greyscale image. The effect is a list of objects containing the contour detail, as well as the dimensions of the image's bounding

 DOI : 10.35291/2454-9150.2021.0129

ISSN : 2454-9150 Vol-07, Issue-01, APR 2021

box. This list is then passed through the pipeline of our algorithm's final step.

C. Text Recognition

We haven't created an individual text recognition algorithm since we're focused on closing the study gap in text detection. Instead, the Tesseract OCR algorithm is used by the machine to identify text in the contours. The technique involves moving each item from the list to Tesseract, which operates on the isolated portion of the image discovered by the bounding box. As a result, the text is succinct and contextually accurate, avoiding the issue that standard OCR approaches have of maintaining the semantic credibility of an image in document format.

V. RESULT

We put our model to the test on a variety of printed and natural invoice images, and it performed admirably, easily outperforming state-of-the-art deep learning-based text detection algorithms. We primarily compared our observations to the east text detection algorithm, which is commonly accepted as the industry norm. We were unable to perform more analysis on more structured text detection benchmarks because they are focused on natural scene images, and generating a result based on comparison with such a dataset does not offer a reliable measure of our system's effectiveness. Figure 3 shows a comparison of the proposed method and the EAST text detection algorithm.

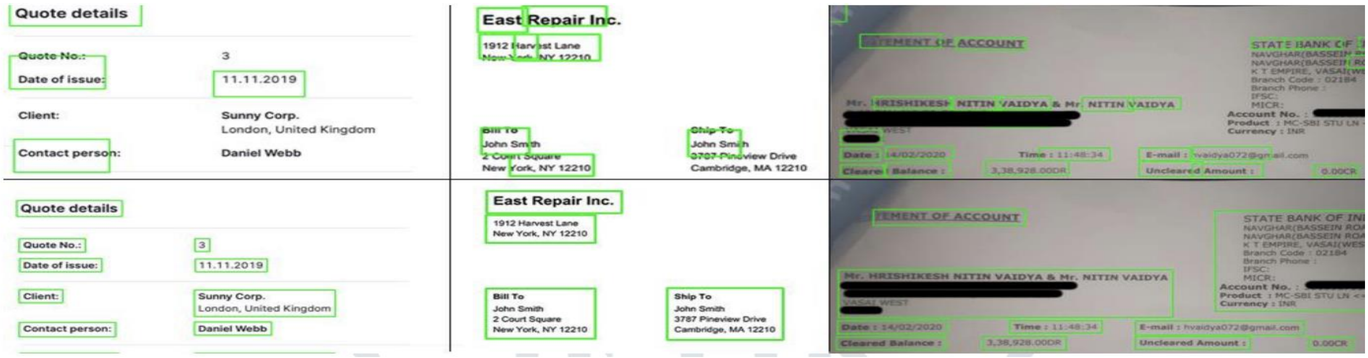

Figure 3. Results of Proposed System vs. Existing System.

VI. CONCLUSION

Throughout its growth, the “Optical Character Recognition using Deep Learning” Method has aimed to maximise productivity while staying efficient and revolutionary. The disadvantages in traditional methods have been minimised. OCR can be used in a wide range of real-time applications. It can be used for a range of items, including workplace automation. This work gives an acceptable solution to the problem. The proposed approach uses image recognition to identify contours in an image, then uses Tesseract OCR to translate each contour into text format individually. We will make a lot of improvements to our new proposed structure to strengthen it. Our first and most significant adjustment will be to allow the user to draw his own contours within the image, as our model is bound to make mistakes when detecting these objects, and in order to avoid data loss, the user should draw his own contours, which the OCR system can detect alongside the contours already imposed by our algorithm. Another change would be to export the data in excel format, making it easy for the administrator to store the information and make changes to the entry fields.

REFERENCES

[1] Zihao Liu, Qiwei Shen, Chun Wang, “Text Detection in Natural Scene Images with Text Line Construction” IEEE International Conference on Information Communication and Signal Processing (ICSP 2018)

[2] Xinyu Zhou, et al. “EAST: An Efficient and Accurate Scene Text Detector” arXiv preprint arXiv:1704.03155(2017)

[3] Tian, Zhi, et al. "Detecting text in natural image with connectionist text proposal network." European Conference on Computer Vision. Springer, Cham, 2016.

[4] Zhang, Zheng, et al. "Multi-oriented text detection with fully convolutional networks." arXiv preprint arXiv:1604.04018(2016).

[5] Cong Yao, Kai Chen, et al. “Real-Time Scene Detection with Differentiable Binarization.”arXiv:1911.08947(2019)

[6] Enze Xie, Gang Yu, et al. “Scene Text Detection with Supervised Pyramid Context Network.” arXiv:1811.08605(2018)

[7] Linje Xing, Zhi Tian, et al. “Convolutional Character Networks.” arXiv:1910.07945(2019)

[8] Ray Smith, “An Overview of the Tesseract OCR Engine” Google Inc.

[9] Tian, Shangxuan, et al. "Text flow: A unified text detection system in natural scene images." Proceedings of the IEEE international conference on computer vision. 2015.

[10] He, Pan, et al. "Reading Scene Text in Deep Convolutional Sequences." AAAI. Vol. 16. 2016.

[11] P. He, W. Huang, T. He, Q. Zhu, Y. Qiao, and X. Li. “Single shot text detector with regional attention.” IEEE International Conference on Computer Vi-sion (ICCV)

[12] B. Shi, X. Bai, and C. Yao. “An end-to-end trainableneural network for image-based sequence recognition and its application to scene text recognition.” IEEE transactions on pattern analysis and machine intelli-gence, 39(11):2298–2304, 2017.

 DOI : 10.35291/2454-9150.2021.0129