A Novel Approach for Invoice Text Detection and Recognition

Hrishikesh Vaidya
BE Student
Department of Computer Engineering
Universal College of Engineering

Aditya Ajgaokar
BE Student
Department of Computer Engineering
Universal College of Engineering

Shounak Daptardar
BE Student
Department of Computer Engineering
Universal College of Engineering

Ankita Kadu
Asst Professor
Department of Computer Engineering
Universal College of Engineering

Abstract—**Text detection has often been a major field of research in computer vision. Previous approaches by the academics have achieved promising results with regard to scene text detection, but these systems usually lack in detecting text that is present in a document format. Traditional OCR engines generally fail when the text in the image doesn't follow a linear format. The proposed system specializes in providing a fast, and accurate text detection system for images that contain systematic and spatial data which is challenging when processed through traditional OCR engines. Our methodical improvements result in efficient text detection through a system that is reliable, fast, and intuitive. It performs well on text that is non-linear in nature and is segmented throughout the image. In comparison to other text detection algorithms such as EAST text detection, our implementation easily outperforms the state-of-the-art text detection techniques. Our proposed system would be highly beneficial for businesses and corporations in handling their financial data.**

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

I. INTRODUCTION

For any kind of financial transaction, an Invoice is generated which contains details of the transaction in a very concise and methodical manner. Corporations usually require to maintain a record of their purchases in their internal database and to perform this action, manual labor is used. This results in a loss of corporate time, and finances. The organization's resources are compromised without any virtual benefit. Modern OCR engines do not consider the fact that the text in an invoice does not follow a linear pattern and is spatially divided based on relevant content.[1] Furthermore, not all invoices share the same structure, so predefining a set of standards for detecting that format is futile. Traditionally, text extraction from an image follows two steps, text detection and text recognition.

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

The proposed system bridges this gap between OCR engines and Invoices by preprocessing the document and detecting text fields that are structured to signify a common element within the invoice or document. These elements would then be individually processed through the Tesseract OCR [8] engine to recognize the text embedded in those elements. This entire process is done through a single pipeline, so that the system remains seamless and as efficient as possible.

II. RELATED WORKS

In recent years, Deep Learning technologies have advanced the efficiency of text detection, which are basically sliding window methods.[9] They all have their enhancements, which mostly involve taking advantage of high-level deep features by using very deep CNN and sharing convolutional mechanisms[10] which have often been implemented to minimize computational costs. Therefore, several FCN-based methods have been proposed and promising results have been obtained in tasks of text detection. Earlier approaches have only focused on one of the two tasks, namely text detection and text recognition. These tasks are then scheduled sequentially to provide the complete optical character recognition. Almost all the previous approaches focus on scene text detection and recognition.[1][2][3][5] Recent text detection techniques were built primarily on general object detectors with different text-specific modifications.[5]

III. Literature Survey

Tian et al. [3] proposed a Connectionist Text Proposal Network (CTPN) to examine the nature of text, and to detect a text example in a sequence of fine-scale text applications. Similarly, Shi et al. developed a linking-segment method which also retrieves a text instance in a sequence, with the ability to identify multi-oriented text. EAST Text Detection [2] was also introduced that explored IOU loss to detect multiple letters within the same parameter that would resemble a word and it produced optimal results. A single-shot text detector (SSTD) [11] has recently been proposed to expand SSD object detector to text detection. SSTD encodes regional attention to text into convolutional features to boost information about text.

Recent work on text recognition inspired by speech recognition commonly casts it into a sequence-to-sequence recognition problem, where recurrent neural networks (RNNs) were employed. He et al. [10] used Convolutional Neural Networks (CNNs) to encode a raw input image into a sequence of deep features, and then an RNN is applied to the sequential encoding and trust mapping features where connectionist temporal classification CTC is applied to produce final results. Shi et al. [12] enhanced these combinations of frameworks by making it trainable end-to-end, with significant performance gain. Recently, the system has been further strengthened by adding different attention mechanisms which can explicitly or implicitly encode more character information.

IV. Proposed Methodology

The proposed system is implemented on synthesized as well as manually captured images of invoices. The system then processes these images by first detecting the areas covered by text, producing bounding boxes besides them and later individually passing each bounding box through the OCR engine which results in segmented text recognition based on categories of text. Training modern Convolutional Neural Networks is an expensive task as it requires state-of-the-art hardware for fast computing. Modern GPUs of the likes of Nvidia and AMD some of which are marketed for the sole purpose of deep learning can compute CNN's optimally, but with a higher cost margin. On top of that, developing a CNN for the sole purpose of OCR has already been heavily researched and there are existing implementations out there that are essentially state of the art OCR systems. In this case, creating our own CNN for OCR purposes would be like reinventing the wheel. The system has instead decided to rely on Google's open source OCR implementation known as Tesseract.

The system also requires the OpenCV image processing library in order to preprocess the image and detect contours that contain an area of segmented text. OpenCV has a lot of inbuilt features and implementations based on popular research papers so it provides better utility in handling the workflow of the system. The proposed system uses a simplistic approach wherein we implement traditional image processing techniques in order to detect regions that contain text in an image. These processing techniques involve a combination of several morphological operations along with other image processing techniques.

A. Image Processing

Image processing comprises several components wherein we process several operations on the image in order to get a contour that provides minimal error. Morphological image processing is a collection of non-linear operations pertaining to the shape or morphology of forms in an image. Morphological operations rely solely on the relative ordering of pixel values, not on their numerical values, and are thus especially suited for manipulating binary images.

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

Morphological operations may also be applied to grey-scale images, such that their light transfer functions are unknown and their absolute pixel values are either of none or minor interest. The proposed system uses morph operations to dilate the image in order to widen the areas of text and strengthen the finalizing contours. Taking everything into consideration, it improves the anchor mechanism to perform well in predicting text components in different levels in text detection task. Note that a text line can be viewed as a series of fine-scale text proposals that can to some degree be treated as a function for object detection. This proposal can include a part of text line and have all the text features and we assume it can work reliably on text detection of multiple sizes and different aspect ratios by detecting a series of text proposals from a text line

B. Text Detection

Text Detection is the most crucial part of the algorithm wherein the result from the preprocessing operation is taken and contours are formed along the main parts of the text. These contours are formed by strategically forming bounding boxes along the regions of the image which is highlighted in the preprocessed greyscale image. The result is a list of objects that contains the contour information including the dimensions of the bounding box on the image. This list is then processed through the last stage in our algorithm's pipeline.

C. Text Recognition

Due to our focus on overcoming the research gap from a text detection perspective we have not developed an independent text recognition algorithm. Instead, the system implements the Tesseract OCR algorithm to recognize text in the contours. The process consists of passing each object from the list to Tesseract, wherein it works on the specific isolated portion of the image which the bounding box has discovered. This results in the text being concise and contextually relevant, eliminating the problem faced by traditional OCR methods where they fail to maintain the structural integrity of an image in document format.

V. RESULT

We tested our model on several printed as well as natural invoice images and overall, it achieved excellent performance that easily outperforms the state-of-the-art deep learning based text detection algorithms. We primarily compared our results with the east text detection algorithm which is widely considered as the state of the art. We could not experiment further on more standardized text detection benchmarks as they are based on natural scene images and producing a result based on comparison with such a dataset would not be an accurate estimation of the effectiveness of our system. The comparison between the proposed system and the EAST text detection algorithm can be seen in Fig.3.

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

VI. CONCLUSION

The "Optical Character Recognition using Deep Learning" System throughout conception, development has aimed at improving the efficiency while being effective and innovative. The drawbacks of conventional systems have been reduced to a minimum. OCR has a wide variety of real time applications. It can be used for many purposes like office automation. This work provides a suitable solution for that problem. The proposed method uses image processing to determine contours in the image and then uses the Tesseract OCR to individually convert those contours into text format We can improve our current proposed system in a wide variety of changes. Our first and most important change will be to give the user the ability to make his own contours inside the image since our model is bound to make some errors while detecting these things and in order to prevent loss of data the user can add his own contours which the OCR system will then detect along with the already placed contours by our algorithm. Another modification will be to output this data in excel format which will make it easier for the user to store this data and modify 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.