Automatic Text Detection from Images: A Deep Learning Approach

S Sireesha

Department of Computer Science Sri Venkateswara University, Tirupati

Abstract— *Our goal is to digitize medical laboratory reports for electronic health record systems using optical character recognition (OCR) techniques. Despite OCR's advancements, challenges persist, especially in medical scenarios. Our approach focuses on segmenting textual information from laboratory report images using deep learning. We designed a concatenation structure for text detection to enhance accuracy, addressing the data-sharing issue among healthcare professionals.*

I. INTRODUCTION

The adoption of electronic health records (EHRs) is crucial for modern medicine, yet complete records are often unavailable due to system issues. Text detection and recognition have gained importance, driven by advancements in computer vision and machine learning. Our focus is digitizing medical laboratory reports for EHRs using optical character recognition (OCR). Despite OCR's progress, challenges persist, especially in medical scenarios. Our deep learning approach addresses these challenges by utilizing a patch-based strategy and a concatenation structure. Results show effective text detection and recognition from medical laboratory reports.

II. LITERATURE SURVEY

2.1 E2E-MLT- An Unconstrained End-To End Method for Multi-Language Scene Text.

M. Buta, Y. Patel, And J. Matas (2018)

In this article, we propose an end-to-end trainable (fully differentiable) method for multi-language scene text localization and recognition. The approach is based on a single fully convolutional network (FCN) with shared layers for both tasks. E2E-MLT is the first published multi-language OCR for scene text. While trained in multi-language setup, E2E-MLT demonstrates competitive performance when compared to other methods trained for English scene text alone. The experiments show that obtaining accurate multi-language multi-script annotations is a challenging problem.

2.2 Textboxes++: A Single-Shot Oriented Scene Text Detector.

M. Liao, B. Shi, And X. Bai (2018).

In this article, we present an end-to-end trainable fast scene text detector, named Textboxes++, which detects arbitrary-oriented scene text with both high accuracy and efficiency in a single network forward pass. No post-processing other than efficient non-maximum suppression is involved. We have evaluated the proposed Textboxes++ on four public data sets. In all experiments, TextBoxes++ outperforms competing methods in terms of text localization accuracy and runtime. More specifically, TextBoxes++ achieves an f-measure of 0.817 at 11.6 frames/s for 1024 × 1024 ICDAR 2015 incidental text images and an f-measure of 0.5591 at 19.8 frames/s for 768 × 768 COCO-Text images. Furthermore, combined with a text recognizer, TextBoxes++ significantly outperforms the state-of-the-art approaches for word spotting and end-to-end text recognition tasks on popular benchmarks.

2.3 Pixellink: Detecting Scene Text Via Instance Segmentation

D. Deng, H. Liu, X. Li (2018)

In this article, the paper demonstrated the PixelLink in novel scene text detection algorithm based on instance segmentation, is proposed. Text instances are first segmented out by linking pixels within the same instance together. Text bounding boxes are then extracted directly from the segmentation result without location regression. Experiments show that, compared with regression-based methods, PixelLink can achieve better or comparable performance on several benchmarks, while requiring many fewer training iterations and less training data.

2.4 Textsnake- A Flexible Representation for Detecting Text of Arbitrary Shapes.

S. Long, J. Ruan, W. Zhang (2018).

In this article, we propose a more flexible representation for scene text, termed as extit {TextSnake}, which is able to effectively represent text instances in horizontal, oriented and curved forms. In TextSnake, a text instance is described as a sequence of ordered, overlapping disks centered at symmetric axes, each of which is associated with potentially variable radius and orientation. Such geometry attributes are estimated via a Fully Convolutional Network (FCN) model. In experiments, the text detector based on TextSnake achieves state-of-the-art performance on Total-Text and SCUT-CTW1500, the two newly published benchmarks with special emphasis on curved text in natural images, as well as the widely-used datasets ICDAR 2015 and MSRA-TD500. Specifically, TextSnake outperforms the baseline on Total-Text by more than extit in F-measure.

2.5 EAST: An Efficient and Accurate Scene Text Detector.

X. Zhou, C. Yao, H.Wen (2017).

In this work, we analyse a simple yet powerful pipeline that yields fast and accurate text detection in natural scenes. The pipeline directly predicts words or text lines of arbitrary orientations and quadrilateral shapes in full images, eliminating unnecessary intermediate steps (e.g., candidate aggregation and word partitioning), with a single neural network. The simplicity of our pipeline allows concentrating efforts on designing loss functions and neural network architecture. Experiments on standard datasets including ICDAR 2015, COCO-Text and MSRA-TD500 demonstrate that the proposed algorithm significantly outperforms state-of-the-art methods in terms of both accuracy and efficiency. On the ICDAR 2015 dataset, the proposed algorithm achieves an F-score of 0.7820 at 13.2fps at 720p resolution.

2.6 Character-Based Handwritten Text Transcription with Attention Networks.

J. Poulos and R. Valle (2017).

In this article, the paper implemented the task of handwritten text transcription with attentional encoder-decoder networks that are trained on sequences of characters. We experiment on lines of text from a popular handwriting database and compare different attention mechanisms for the decoder. The model trained with softmax attention achieves the lowest test error, outperforming several other RNN-based models. Softmax attention is able to learn a linear alignment between image pixels and target characters whereas the alignment generated by sigmoid attention is linear but much less precise. When no function is used to obtain attention weights, the model performs poorly because it lacks a precise alignment between the source and text output.

2.7 Focusing Attention- Towards Accurate Text Recognition In Natural Images.

Z. Cheng, F. Bai, Y. Xu (2017).

In this article, we propose the FAN (Focusing Attention Network) method that employs a focusing attention mechanism to automatically draw back the drifted attention. FAN consists of two major components: an attention network (AN) that is responsible for recognizing character targets as in the existing methods, and a focusing network (FN) that is responsible for adjusting attention by evaluating whether AN pays attention properly on the target areas in the images. Furthermore, different from the existing methods, we adopt a ResNet-based network to enrich deep representations of scene text images. Extensive experiments on various benchmarks, including the IIIT5k, SVT and ICDAR datasets, show that the FAN method substantially outperforms the existing methods.

2.8 Scene Text Detection And Recognition: Recent Advances And Future Trends.

Y. Zhu, C. Yao, And X. Bai (2016).

In this article, the paper implemented the three-fold: 1) introduce up-to-date works, 2) identify state-of-the-art algorithms, and 3) predict potential research directions in the future. The rich and precise information embodied in text is very useful in a wide range of vision-based applications, therefore text detection and recognition in natural scenes have become important and active research topics in computer vision and document analysis. Moreover, this paper provides comprehensive links to publicly available resources, including benchmark datasets, source codes, and online demos. In summary, this literature review can serve as a good reference for researchers in the areas of scene text detection and recognition.

2.9 Text Detection and Recognition In Imagery: A Survey.

Q. Ye And D. Doermann (2015)

In this article, we are analyzing, compares, and contrasts technical challenges, methods, and the performance of text detection and recognition research in color imagery. It summarizes the fundamental problems and enumerates factors that should be considered when addressing these problems. Existing techniques are categorized as either stepwise or integrated and sub-problems are highlighted including text localization, verification, segmentation and recognition. Special issues associated with the enhancement of degraded text and the processing of video text, multi-oriented, perspectively distorted and multilingual text are also addressed. The categories and sub-categories of text are illustrated, benchmark datasets are enumerated, and the performance of the most representative approaches is compared. This review provides a fundamental comparison and analysis of the remaining problems in the field.

2.10 Robust Scene Text Detection with Convolution Neural Network Induced MSER Trees.

W. Huang, Y. Qiao (2014).

In this article, we propose a novel framework to tackle this problem by leveraging the high capability of convolutional neural network (CNN). In contrast to recent methods using a set of low-level heuristic features, the CNN network is capable of learning high-level features to robustly identify text components from text-like outliers (e.g. bikes, windows, or leaves). Our approach takes advantages of both MSERs and sliding-window based methods. The MSERs operator dramatically reduces the number of windows scanned and enhances detection of the low-quality texts. While the sliding-window with CNN is applied to correctly separate the connections of multiple characters in components. The proposed system achieved strong robustness against a number of extreme text variations and serious real-world problems. It was evaluated on the ICDAR 2011 benchmark dataset, and achieved over 78% in F-measure, which is significantly higher than previous methods.

Problem Definition: The transition to electronic health records (EHRs) is crucial for modern medicine, yet complete records are often lacking due to system issues. Medical laboratory reports play a vital role in patient assessment, diagnosis, and monitoring. Our focus is on digitizing these reports for EHR systems, using optical character recognition (OCR) techniques. Despite OCR's progress, challenges persist, especially in diverse scenarios and with lower-quality data. In our method, text recognition is based on CRNN and enhanced with a concatenation structure, enabling direct output of text sequences for detected textual objects..

Drawbacks

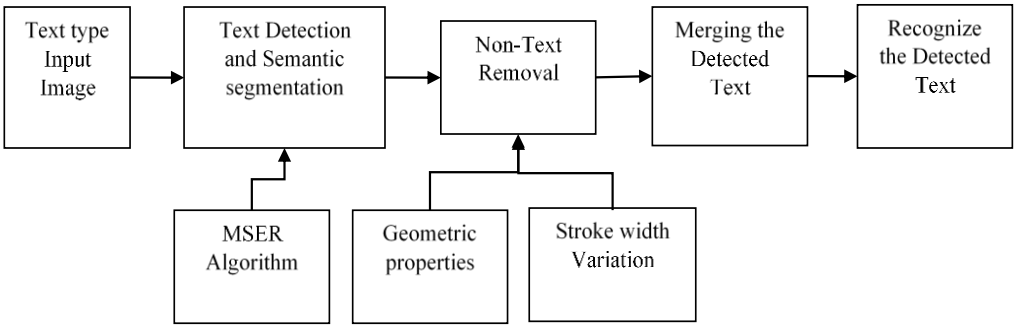
- Not Effectively Detected
- Not Accurate
- Less Performance
- Inefficiency

III. PROPOSED WORK

The transition to electronic health records (EHRs) is crucial for modern medicine, but complete records are often unavailable due to system issues. This paper proposes a deep-learning approach for segmenting textual information from laboratory report images, aiding data sharing among physicians. Our method employs a concatenation structure for text detection, improving multi-lingual text recognition accuracy. In real-time applications, such as alerting drivers about road signs, our approach detects and recognizes text automatically from captured video, enhancing optical character recognition (OCR) tasks

Advantages

- Accurate.
- More Text can detect.
- Efficient.
- Better Performance.



3.1 Input Image

A collection of data is called dataset. Deep learning requires massive amount of training dataset as classification accuracy of deep learning classifier is largely dependent on the quality and size of the dataset, however, unavailability of dataset is one the biggest barrier in the success of deep learning. Here, an input image can be obtained in the text format image. Text can appear differently in the images.

3.2 Text Detection and Segmentation

The text detection function determines text presence through localization and verification, providing precise text instance images for recognition. Text segmentation, challenging yet crucial, involves text line and character separation. Semantic Segmentation is employed for text detection. The automated algorithm detects numerous text region candidates and progressively filters out non-text regions using the MSER algorithm, known for its stability in detecting text due to consistent color and high contrast.

3.3 Nontext Removal

A non-text removal is used to remove the unwanted text from the image. A non-text removal approaches based on the Geometric Properties and stroke width Variation.

3.3.1 Based on the Geometric Properties:

Although the MSER algorithm picks out most of the text, it also detects many other stable regions in the image that are not text. You can use a rule-based approach to remove non-text regions. Geometric properties of text can be used to filter out non-text regions using simple thresholds. Alternatively, you can use a machine learning approach to train a text vs. non-text classifier. Typically, a combination of the two approaches produces better results of the of the system. This example uses a simple rule-based approach to filter non-text regions based on geometric properties.

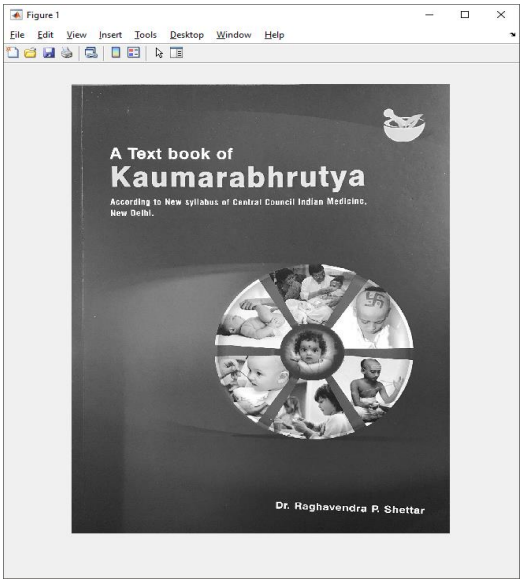
3.3.2 Based on the stroke width Variation:

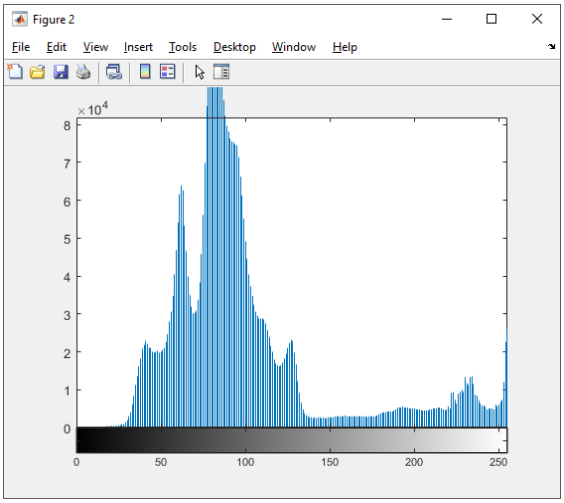
Another common metric used to discriminate between text and non-text is stroke width. Stroke width is a measure of the width of the curves and lines that make up a character. Text regions tend to have little stroke width variation, whereas non-text regions tend to have larger variations. The stroke width can be used to remove non-text regions, estimate the stroke width of one of the detected MSER regions. By using a distance transform and binary thinning operation is used in this model of the system.

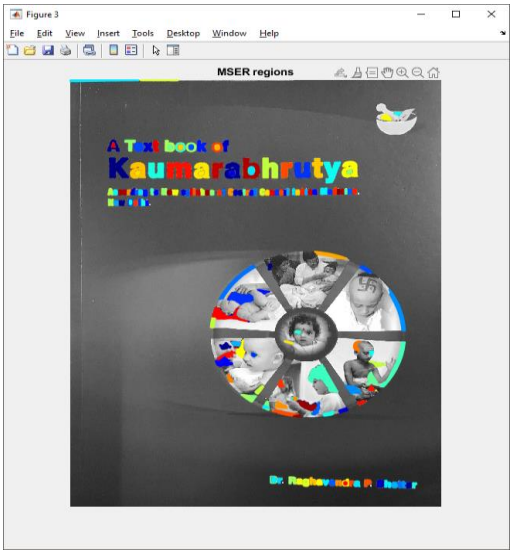
3.4 Merging and Recognizing the Detected Text

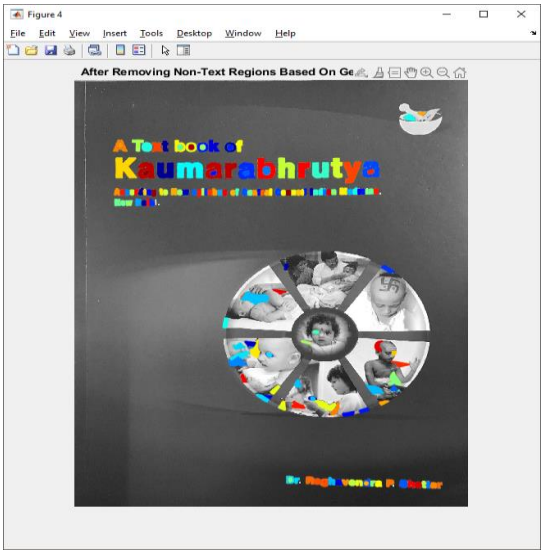
The detection results, initially comprising individual characters, need to be merged into words or text lines for recognition tasks like OCR. One method involves finding neighboring text regions and forming bounding boxes around them. Overlapping bounding boxes of neighboring text regions are then merged to create single bounding boxes around words or text lines. Using a graph, connected text regions with non-zero overlap ratios are identified. After text detection, OCR is applied to recognize text within each bounding box. To rectify noisy OCR output, MSER is used. Finally, the predicted text from the images is outputted.

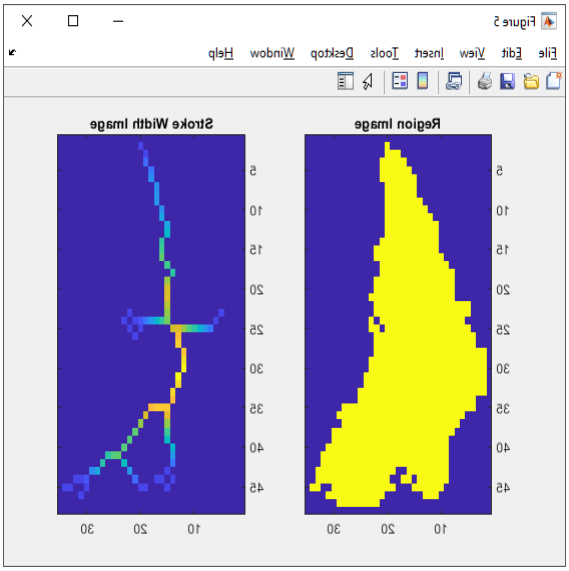
IV. IMPLEMENTATION

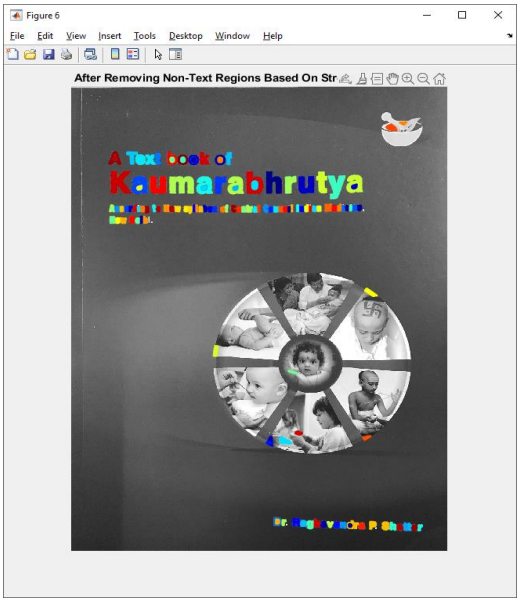


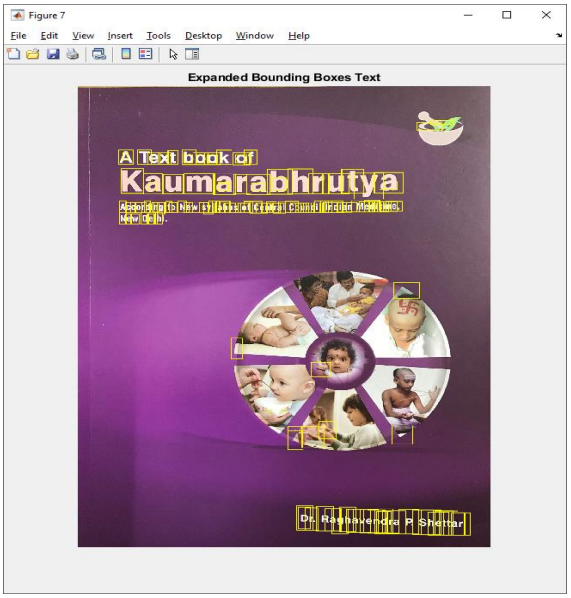


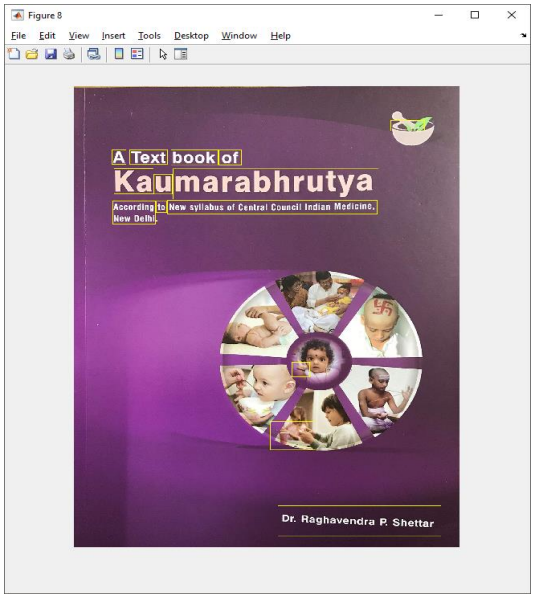


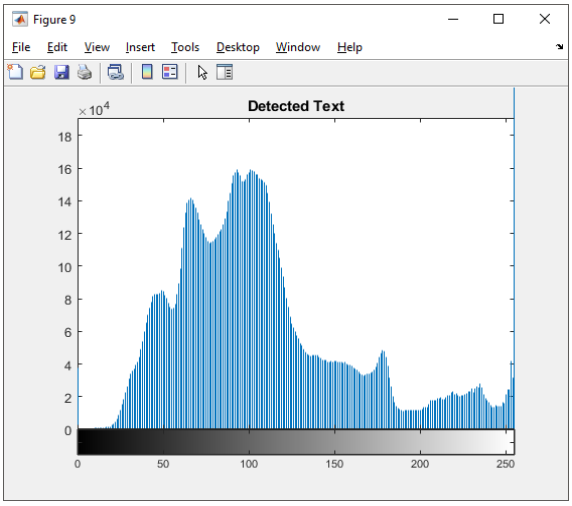












ans =

'Ka marabhrutya

Dr. Raghavendra P. Shettar

book of

According

New Delhi

New syllabus of central council Indian Medicine,

I0

Text

V. CONCLUSION

This paper introduces a deep learning method for text detection and recognition in medical laboratory report images. Using a patch-based training approach, a detector outputs bounding boxes containing texts, while a concatenation structure in a recognizer identifies recognized texts from these boxes. Our text detection module, enhanced with a patch-based strategy, achieves better accuracy. Experimental results show the effectiveness of the concatenation structure in combining features for recognition. The approach handles images with varying resolutions well. Conclusion: Automatic text detection in natural images has many applications, including reducing manual transcription costs in healthcare services in developing countries. Structured health records from document images can enhance medical data mining for improved healthcare services in the future.

REFERENCES

[1] Khan, Tauseef & Sarkar, Ram & Mollah, Ayatullah. (2021). Deep learning approaches to scene text detection: a comprehensive review. Artificial Intelligence Review. 54. 1-60. 10.1007/s10462-020-09930-6.

[2] Amritha S Nadarajan, Thamizharasi A, “A Survey on Text Detection in Natural Images", International Journal of Engineering Development and Research (IJEDR), ISSN: 2321-9939, Volume.6, Issue 1, pp.60-66, January 2018.

[3] Tridib Chakraborty et al, (2017), Text recognition using image processing, International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017, 765-768

[4] Karaoglu, S., Tao, R., van Gemert, J. C., & Gevers, T. (2017). Con-Text: Text Detection for Fine-Grained Object Classification. IEEE Transactions on Image Processing, 26(8), 3965-3980.

[5] Guan, L., & Chu, J. (2017, June). Natural scene text detection based on SWT, MSER and candidate classification. In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on (pp. 26-30). IEEE.

[6] Zhu, Q. H., Zhu, R., Li, N., & Yang, Y. B. (2017, October). Deep metric learning for scene text detection. In Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on (pp. 1025-1029). IEEE.

[7] Shi, B., Bai, X., & Belongie, S. (2017). Detecting oriented text in natural images by linking segments. ArXiv preprint arXiv: 1703.06520.

[8] Zhong, Z., Jin, L., & Huang, S. (2017, March). Deeptext: A new approach for text proposal generation and text detection in natural images. In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 1208-1212). IEEE.