Retraction

Retraction: Assistant for the guest with visually impaired using Deep Learning (J. Phys.: Conf. Ser. 1916 012063)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Assistant for the guest with visually impaired using Deep Learning

Priya A¹, Shalini M¹, Suganti T¹, Swetha M¹

¹Department Of Computer Science and Engineering, Sri Krishna College Of Engineering and Technology, Coimbatore, India.
17eucs141@skcte.ac.in

Abstract. The lack of Braille resources in this advanced world has tied the hands of visually impaired people from soaring up. This paper takes those guests into concern and presents a solution that helps every individual especially the blinds in reading books and text from real time images. This solution converts the text obtained from text document and real world entities to aural output which lends a hand in reading text. The main idea is to build a software using a novel methodology where OCR engine receives input images and get converted into intermediate textual output that is given to Google Trans to get the audio output via earphones.

Keywords: Image, Optical Character Recognition (OCR), Tesseract, OpenCV East Detector, LSTM

1. Introduction

The major impediment in the life of Visually impaired is to acquire resources that are compatible for them to read. When they visit public places like shops, restaurants, hospitals they have the need to depend on others. Hence this acts as a stumbling block in their lives. In order to overcome these drawbacks, many evolved to make their lives manageable using various technologies and methodologies. Initially many implementations were discovered which were based on technologies that generate the required text into its Braille equivalent. Tsukuda et al came up with a solution called Braille printer which in turn provide Braille characters for the given input text. It faced certain sort of troublesomeness like time frame, cost, unmanageability in the case of larger texts etc. And thus in this case, this was not able to expect to produce an instantaneous solution and failed to give analogous output for all inputs. Eventually this leads to the rise of other forms of technologies from the feedbacks received.

After Braille auditory design of resources were created and attempted in reading. Books in the form of Cassettes and tapes broke down from continuation spot for the readers when they start reading. It has to be followed in an order and one cannot get or divert back to the continuation spot of the text. It was difficult to hear the text iteratively. DAISY, Digital Accessible Information System is designed to be a complete substitute for audio of print materials and was designed specifically for use by people with blindness, dyslexia and impaired vision. Thus this was able to give a lively reading experience.
D’Albe invented Optophone which was one among the aiding devices for visually impaired that uses musical chords which was able to convert the visual representation of words to phonic output. Goldish and Taylor invented Optical to Tactile Converter (Optacon) where vibrotactile signals are used. Here, while reading the user would move their camera held by hand line by line using the Optacon. This was difficult to orient the text properly. Then an advance in computation led to the next form of innovation. Optical Character Recognition (OCR) was used which extract texts from the image. OCR extract texts from the given input image using corresponding algorithm. Then the evolution of smart phones along with the blooming techniques in OCR and Speech Synthesis improvised and led to the invention of mobile aided reading applications like KNFB Reader. This application gets the input text from the photo captured by camera and produced acoustic output. Alike KNFB, tools like Blind sights Text Detective came into use. The requirement is that the text should be captured from an appropriate for proper and clear input so that it can produce optimum results. Thus such mobile software require correct and proper alignment, lighting and correct focus points that tends to produce optimum output and would help in extracting clustered text from given image. Thus relevant information may not be provided by this kind of software that the user intent to read.

Therefore from the above literatures, it is quite difficult to identify clustered text labeling for colossal datasets with different font styles. Thus a limitation in capability that the information cannot be retrieved has been concluded through these methods. Hence a deep learning model and algorithm that provides optimum results with cost effective methodologies has been imposed in this paper.

2. Literature Survey

Later on, the grams algorithm was utilized around the unique words that match. [1] in "An Expeditious Alignment Scheme for Automatic OCR Evaluation of Books" aims at the precision of the OCR on scanned books. This involved finding distinct words utilizing distance-based algorithms.

A new separation and word spotting algorithms were acclimated to recognize handwritten Arabic words. [2] in "Segmentation-Predicated And Segmentation-Free Methods for Spotting Handwritten Arabic Words" marginally integrates ambiguity into a complete Arabic word; that is, the amount of words in observe that are distinct from one another solely by the presence of characters with the same kindred base shape, however having or not having associated dots isn't immensely colossal.

"Multilingual OCR for Indic Scripts" [3 -5] optically canvassed that, it is even with homogeneous architecture, performance in languages of Indian is arduous compared to English. In the latter case, the data required for RNN training would be less and the output space would be more minute. In this approach, the script is identified at a word level, afore the recognition of the word demonstrated for English language and twelve Indian languages.

"Matching Ottoman Words: An image retrieval approach to historical document indexing" [6-8] was about that the characteristics of Ottoman documents needed an automatic transcription, but the character apperception predicated system was not given any efficient and copacetic results. This study faces the quandary in image retrieval and proposes a solution that was predicated on image matching techniques because of containing consequential images and signatures in Ottoman documents. Additionally, it has to face a drawback of causing most of these Ottoman documents to be inaccessible because many documents are in manuscript and manual transcription format. Additionally, the time taken for the indexing of these Ottoman documents and text was long. So, it leads to an inaccessible state.
"Handwritten Word Spotting with Redressed Attributes" [10-13] suggested that by utilizing an attributes-based approach that leads to a low dimensional, it has a fine-tuned-length representation and cumulated representation of word images and strings, so it is expeditious to compare and compute. It allows us to perform queries indistinctly like a query-by string. The query is an image and string. A calibration scheme was proposed to ameliorate challenging dataset results by Canonical Correlation Analysis. It was tested on two public datasets exhibiting state-of-the-art results. Withal it had a disadvantage of words that were labeled were not utilized in training.

3. Proposed Methodology

The paper focalize on software routine that supports every individual especially the partly or wholly blind in going-over printed version and text in real time images. This work consists of different algorithms and modules namely

- Image Acquisition
- Text recognition
- Text extraction
- Text to audio output

```
3.1 Image Acquisition

Here, the image acquisition takes place with the component camera where images of interest are given as input.

3.2 Image Preprocessing

Once the image is focused under camera, the dimensions of the image are grabbed. Later image captured is amended to the dimension 256 x 256 with the aim that the attesting routine will deed upon evenly classified shots. This gives homogeneous outcome for every input. The re-sized shot or images is then loaded into trained East Text Detector.
```

![Figure 1. Architectural Diagram](diagram.png)
3.3 Text Recognition

A deep learning OpenCV model is used which is a novel architecture and an instruction pattern that is skilled to recognize all real-time images. This obtains up to the minute text detection accuracy. Coming to “EAST” that is an Efficient and Accurate Scene Text detection pipeline which is dynamic to recognize subject matter even though when it’s fuzzy, reflective or even partly concealed using this pre-trained networks. Since they are trained in advance, there is no need to take up much time for the networks to be trained. Instead we can go further and utilize the outcome to make decisions within our application.

Poor OCR is provided when there is no proper orientation with the documents and pics. So now, we'll adjust the orientation of the input image to ensure better OCR results using ROI (Region Of Interest).

Here, the timing of text recognition and the boundaries of the image are found.

3.4 Text Extraction

Once text ROIs are extracted, it is provided to Tesseract which helps in building an entire pipeline of OpenCV OCR. This Tesseract has capability to recognize many languages after getting trained. This OCR engine employs an LSTM (Long Short-Term Memory) network by its own which lends hand in providing optimum output. LSTM is a kind of Recurrent Neural Network (RNN) that is used to maintain the accuracy in text extraction. The text extracted is stored in a result variable. Lastly the result is written in a file.

3.5 Text to Speech

The intermediate result (text) that is stored in a file is converted into audio using Google Trans which is free and this python library is implemented using Google Translate API.

4. Experiment and Result Analysis

Thus the input image is the sequence of information which is given to the processor. The pre-processing of image removes distortions and helps in enhancing important features for further processing. OCR software program frequently "pre-processes" the photographs to improve the possibilities of a hit popularity.

The text is recognized using OpenCV EAST text detector. In order to recognize the text, first text region is grabbed using Region Of Interest (ROI) This ROI provides coordinates of text. The intermediate result is passed to Tesseract which employs LSTM that offers optimum result in recognition. The result of LSTM provides the actual OCR results. Finally the fetched text from the input image is read out through the aural as an outcome.

5. Conclusion

This project is implemented using Python. The proposed assistive software presents a solution that helps every individual especially the blinds in reading books and text from real time images Figure 1, 2, 3 and 4.
Figure 2. Input Image
This will be the input image that will be captured by the user.

Figure 3. Recognising the text from the input image
Here OCR first works by pre-processing the input digital image into smallest component parts with layout analysis to find the word blocks. Other features like lines photographs are recognised and ignored. The word that has been recognized is compared with the OCR engines’s large dictionary for checking the existence.

Figure 4. Printing the text to the terminal
The text in the above output terminal is converted into audio.
References

[1] H. AlSaid, L. AlKhatib, A. AlOraidh, S. AlHaidar and A. Bashar, Deep Learning Assisted Smart Glasses as Educational Aid for Visually Challenged Students, 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923044.

[2] D. Velmurugan, M. S. Sonam, S. Umanaheswari, S. Parthasarathy, K. R. Arun (2016). A Smart Reader for Visually Impaired People Using Raspberry PI. International Journal of Engineering Science and Computing IJESC Volume 6 Issue No. 3.

[3] Silvio Ferreira, Céline Thillou, Bernard Gosselin. From Picture to Speech: An Innovative Application for Embedded Environment. Faculté Polytechnique de Mons, Laboratoire de Théorie des Circuits et Traitement du Signal Bâtiment Multitel - Initialis, 1, avenue Copernic, 7000, Mons, Belgium.

[4] S. Venkateswarlu, D. B. K. Kamesh, J. K. R. Sastry, Radhika Rani (2016) Text to Speech Conversion. Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102967, October 2016.

[5] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770–778.

[6] W. Lu, H. Sun, J. Chu, X. Huang and J. Yu, A Novel Approach for Video Text Detection and Recognition Based on a Corner Response Feature Map and Transferred Deep Convolutional Neural Network, in IEEE Access, vol. 6, pp. 40198-40211, 2018, doi: 10.1109/ACCESS.2018.2851942.

[7] P. Jing, Y. Su, L. Nie, and H. Gu, Predicting image memorability through adaptive transfer learning from external sources, IEEE Trans. Multimedia, vol. 19, no. 5, pp. 1050–1062, May 2017.

[8] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[9] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. 27, no. 4, p. 273, 2018.

[10] Nagaraja L, Nagarjun R S, Nishanth M Anand, Nithin D, Veena S Murthy (2015). Vision based Text Recognition using Raspberry Pi. International Journal of Computer Applications (0975 – 8887) National Conference on Power Systems & Industrial Automation.

[11] Poonam S. Shetake, S. A. Patil, P. M. Jadhav (2014) Review of text to speech conversion methods.

[12] S. Venkateswarlu, D. B. K. Kamesh, J. K. R. Sastry, Radhika Rani (2016) Text to Speech Conversion. Indian Journal of Science and Technology, Vol 9 (38), DOI: 10.17485/ijst/2016/v9i38/102967, October 2016.

[13] Y. Tsukuda, I. Goto, Braille Printer, US5193921A,(1991)