Text extraction from Images using Convolutional Neural Network

Bharati V[1], Sudarshan Rao M [2], Aditi J[3], S G Aditya Bharadwaj[4], S Srividhya[5]

[1,2,3,4]B.E Student, Dept. of Information Science and Engineering, BNM Institute Of Technology, Bengaluru-560070(Karnataka).
[5]Assistant Professor, Dept. of Information Science and Engineering, BNM Institute Of Technology, Bengaluru-560070(Karnataka).

---***---

Abstract - *In this paper, we present a text extraction model that is designed to process the images that areuploaded by the user. Images along with texts have become one of the common ways to exchange information; hence understanding these images plays an important role. We present efficient text detection and extraction models along with search.*

Key Words: Text detection, text recognition, CNN, Text Extraction, Pre-Processing.

1. Introduction

Text detection [2] and extraction is used to get the extracted text in a document using the state-of-the-art algorithms such as Convolutional neural networks and the techniques that follow it. It is going to be very helpful for those who are in data entry department who can get the content of some photos of the bills and invoices directly on their screens rather than typing it out manually. Using the current technology to solve such problems in the real world with system possible solutions is one of the greatest goals of the project.

To get an accurate and relevant search results when one searches for a text in an image. This is to enable people in the industry, to directly get the image by searching for a keyword in that image.

1.1 Innovation Presented

- Browse for the image that contains text.
- Extracting text from the image.
- Implementing the search technique   to identify keyword in the text.
- The text that is extracted can also be stored in a document, which is in an editable format.
- The system is trained efficiently with all combination of inputs using the datasets.

2. Text Extraction Model

This model is done in two steps: detection and recognition. [1] First, we detect those regions in the image potentially containing text. In the second step we perform text recognition, where, for each of the detected regions, a CNN is used to recognize and transcribe the word in the region to detect the text. It represents an image into a convoluted feature map. This map is taken as input producing bounding boxes that contain text [5]. In the last stage, we extract the text.

2.1 Comparison Of Different Technologies

Used For Text Extraction

- Region based Method: Region-based method uses the properties of the color or gray scale in the text region or their differences to the corresponding properties of the background. They are based on the fact that there is very little variation of color within text and this color is sufficiently distinct from it's immediate background. Text can be obtained by thresholding the image at intensity level between the text color and that of its immediate background. This method is not robust to complex background and is further divided into two sub-approaches: connected component (CC) and edge based.

- CC based Method: CC-based methods use a bottom- up approach by grouping small components into successively larger components until all regions are identified in the image. A geometrical analysis is required to merge the text components using the spatial arrangement of those components so as to filter out non-text components and

the boundaries of the text regions are marked. This method locate locates text quickly but fails for complex background.

- Edge based Method: Edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Edge based method is focused on high contrast between the text and the background. The three distinguishing characteristics of text embedded in images that can be used for detecting text are edge strength, density and the orientation variance. Edge based text extraction algorithm is a general-purpose method, which can quickly and effectively localize and extract the text from both document and indoor/ outdoor images. This method is not robust for handling large size text.

- Texture based Method: This method uses the fact that text in images has discrete textural properties that distinguish them from the background. The techniques based on Gabor filters, Wavelet, Fast Fourier transform (FFT), spatial variance, etc. are used to detect the textual properties of the text region in the image . This method is able to detect the text in the complex background. The only drawback of this method is large computational complexity in texture classification stage.

- Morphological based Method: Mathematical morphology is a topological and geometrical based method for image analysis. Morphological feature extraction techniques have been efficiently applied to character recognition and document analysis. It is used to extract important text contrast features from the processed images. These features are invariant against various geometrical image changes like translation, rotation, and scaling. Even after the lightning condition or text color is changed, the feature still can be maintained. This method works robustly under different image alterations.

3. Pre- Processing

In this system, there are 4 different types of filters used to pre-process the image:

1. Binarize
2. Median blur
3. Scale up
4. De-skew

A combination of these filters is used in order to get the highest accuracy in extracting all the text present in the image.

4.   Text Recognition Model

The text recognition model is a CNN based on the ResNet18 [4] architecture. To train [3] the model, cast it as a sequence prediction problem, where the input is the image containing the text to be recognized and the output is the sequence of characters in the word image. Use the connectionist temporal classification (CTC) loss to train the sequence model. Casting the issue as one of sequence prediction allows the system to recognize words of arbitrary length and to recognize out-of-vocabulary words (i.e., words that weren't seen during training).
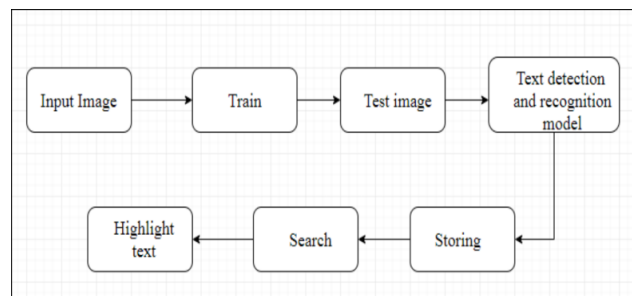

Fig -1: Architecture of the proposed system

4.1 Technologies used for Text Recognition

Convolution Neural Network

CNN is a class of deep, feed-forward artificial neural networks (where connections between nodes do not form a cycle) & use a variation of multilayer perceptrons designed to require minimal pre-processing. These are inspired by animal visual cortex. Convolutional neural network are used to find patterns in an image. We do that by convoluting over an image and looking for patterns. In the first few layers of CNNs the network can identify lines and corners, but we can then pass these patterns down through our neural net and start recognizing more complex features as we get deeper.

Recurrent Neural network

A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit
dynamic temporal behavior for a time sequence. RNN is a
sequence of neural network blocks that are linked to each other's like a chain. Each one is passing a message to a successor.

5. Searching Model

There are two stages that are implemented in search:

- Identifying the keyword.
- Highlighting the keyword in the text once found.

First take an input from the text field, For example: "Good Evening". This is converted into a JavaScript regular expression- "/\b (Good| Evening) \b/". This regular expression will match any of the entered words where they appear in the content area of the page. The apply() method is called, it generates a regular expression from the keywords, clears any existing highlights on the page, and then calls the highlightWords() method passing a reference to the selected start node.
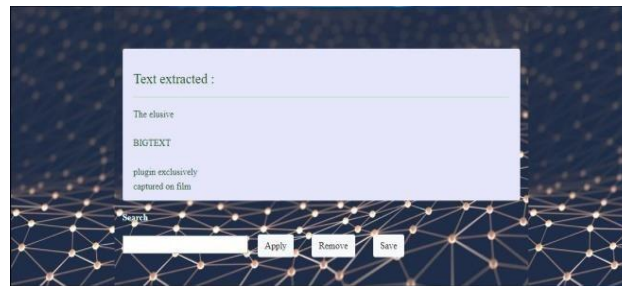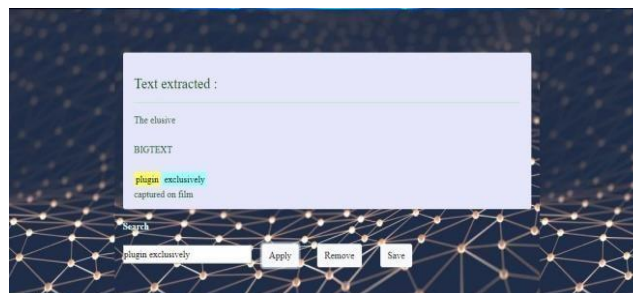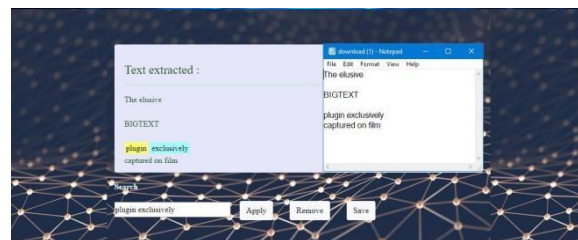
6. Output

Fig -2: Browse the Image

Fig -3: Pre-process the image

Fig -4: Text extracted

Fig -5: Search for keyword

Fig -6: Storing the text in Document

7. Conclusions

Today most of the information is available either on paper or in the form of photographs. The current technology is restricted to extracting text against clean backgrounds. Thus, there is a need for a system to extract text from general backgrounds. Text Extraction and recognition in Images has become a potential application in many fields like [6] Image indexing, Robotics, Intelligent transport systems etc. For example capturing license plate information through a video camera and extracting license number in traffic signals. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the hitch of automatic text extraction extremely challenging.

The current system can extract the text from images and perform search action for a particular keyword and also store the extracted text in a document which is in an editable format.

8. Future Enhancement

Future work of the project may include developing an application for the smart phones and making it available in cross platform and also improving the user interface. Make it multi-lingual text extraction. In Military, the code maybe written in such a way that, white text is on white background and extracting such text might be critical at the situation. Detecting watermarks.

References

1) Fedor Borisyuk, Albert Gordo and Viswanath Sivakumar, "Rosetta: Large Scale System for Text Detection and Recognition in Images", KDD 2018, August 19-23, 2018, London, United Kingdom.

2) Manolis Delakis and Christophe Garcia, "Text Detection with Convolutional Neural Networks", VISAPP 2008 - International Conference on Computer Vision Theory and Applications.

3) Andreas Veit, Tomas Mater, Lukas Neumann, Jirı Mata and Serge Belongie, "COCO-Text: Dataset And Benchmark For Text Detection And Recognition In Natural Images", arXiv:1601.07140v2 [cs.CV] 19 Jun 2016.

4) Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun,"Deep Residual Learning for image recognition", IEEE conference on computer vision and pattern recognition, 2016, page- 770-778.

5) Uma B Karanje and Rahul Dagade,"Survey On Text Detection, Segmentation and Recognition From Natural Scene Image", International Journal of Computer Applications, 2014.

6) Shubhrita Tiwari Shailendra Singh Kathait," Application of Image Processing and Convolution Networks in Intelligent Character Recognition for Digitized Forms Processing", Int. J. Comput. App, 2018.