International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Image to Text conversion in Foreign Language using Document Image Processing Technique

Rishabh Jain, Satyam Gupta, Shahzeb Rizvi, Nitin Arora

Abstract: ***Every country has their own native languages such France has French, Japan has Japanese, India has Hindi and other local languages like Gujrati, Marathi, Telugu, Tamil, Bengali, etc. When a person who does not know the English Language travel to a country whose local language is English will face many problems like understanding road sign texts, shop's name, Instructions Boards etc. So, in this paper we are proposing a model with the help of which user can take snap of the scene containing text of which he/she want to translate it, upload that photo, choose the language and the software will provide us the output in text of User's native language. This model will use Convolutional Neural Network (CNN) to identify the characters, Efficient and Accurate Scene Text Detector (EAST) for detecting the text in the image, Digital Image Processing methods are used to segment the text in the detected text, googletrans for translating the text to other language and Tkinter for creating Graphical user Interface (GUI) for software.***

Keywords **: *GUI, CNN, EAST Detector, Googletrans***

I. INTRODUCTION

Early years, Human Translators are used to set the communication between different country people. Having human translator also have various cons such as turnaround time is longer, a good income should be given to human translators, human translators are restricted to certain number of languages, etc. Mobile devices can be used to provide machine translation services from a given image. The service is commonly termed as Image translation for mobile devices. In this whole process, the user will take or provide photograph of some written text and the text would be extracted from the image and translated into a language preferred by them. Optical Character Recognition (OCR) is a term that is used for mechanical or electrical conversion of images of written or printed text into text that are of a specified language.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Rishabh Jain*, School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India. Email: rishabhking05@gmail.com
Satyam Gupta, School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India. Email: satyam.g16@gmail.com
Shahzeb Rizvi, School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India. Email: shahzebr74@gmail.com
Nitin Arora, School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India. Email: narora@ddn.upes.ac.in

It can be used for scanning documents or understanding billboards and signs. It is a run of the mill procedure of digitizing composed messages so they'll be electronically adjusted, looked, put away more minimally, showed on-line, and utilized in machine procedures. OCR is a field that could be analyzed more in various data science and image processing fields like artificial intelligence and computer vision.

Efficient and Accurate Scene Text (EAST) detector is a powerful tool for text detection in a surrounding area. It is mostly used for text detection and segmentation in an image. It is a pre-trained model which provided an accuracy of 81 F1-score on ICDAR 2015 and 2013. Character detection and segmentation can be done in many different ways but we have done it with a strip-by-strip detection according to our target images.

Hand-written text cannot be done by this method as the characters are interlinked most of the time. Googletrans is a library that can be installed in python programming language for text or sentence translation. It uses Google Neural Machine Translation (GNMT) to translate.

It can detect the input language and can convert into any language input by the user. In this article we would be talking about Convolutional neural network and using it for character labeling with various proposed algorithms and methods of image processing for image segmentation. Literature reviews are the brief of the references provide in the last section. We would also share our results and conclusion after trying various ways to come up with better OCR model. There are references in the last section of this article that helped us in completing and understanding the project.

II. LITERATURE REVIEW

In 2017 X. Zhou et al.[1], gave a methodology to detect the text in a natural scene image with the help of 2 output layers. The first layer is used for output probabilities and the second layer is used to find the bounding box coordinates. The limitation of this algorithm is that text is detected only on the basis of confidence score. If we set confidence score high then we face problem in detecting the whole text and if decrease it then it will detect those regions that do not contains text.

Albawi et al. [2] described the overall understanding of neural network and use of Convolution Neural Network on various layers for image feature extraction and converting them into feature vectors for further be classified by the Artificial Neural Network Layers. The layers used in CNN are [12]

Convolutional Layer for crossing the image with various feature detectors. ReLu Layer for converting the image into favorable pixel values Max Pooling Layer to reduce the image size for further processing Flattening Layer for converting the images to vector values Full-Connection – connecting the flattening layer to the ANN layers. This method of classification can only be used for a single character recognition and detection and it cannot be used for text recognition as it would make it a very complex process.

Shin et al. [3] evaluated CNN (Convolutional Neutral Network) in three important factors Dataset characteristics, Architecture and Transfer Learning. They evaluated CNN performance on two different computer-aided diagnosis applications Thoraco-abdominal lymph node detection and interstitial lung disease classification.

Tsung-Yu Ling et al. [4] came up with a new architecture for CNN model for fine-grained categorization. The model consists of two feature extractors who is operated and pooled to obtain an image descriptor. The data structure of their model used two stream of CNN model which were crossed down with a bilinear vector and then softmax function was applied for categorization of the output. The accuracy of their datasets where up to 84.1%.

Baohua Sun et al. [5] proposed a two-dimensional word embedding for tabular data classification which is the most common data in industry. Previously used embedding were one-dimensional. The features are first fine-tuned in two-dimensional CNN models. The result where great for both small and large data sets.

Rebort Dürichen et al. [6] came up with the bases of binary output prediction in deep convolutional neural networks. They presented the concept of binary input layer.

Yann LeCun et al. [7] proposed article tells us about the use of Gradient-Based Learning technique and how it is used in multilayer neural networks with back-propagation like CNN. It goes in detail about the mathematics applied in these models and how it is effective.

Jainxin Wu et al. [8] provided an 31 page read of CNN introduction containing vectorization, chain rule, architecture, stochastic gradient descent, layers i.e., input layer, output layer, ReLu layer, convolution layer, pooling layer, etc. It is full study of CNN model.

Rob DiPietro et al. [9] article on cross-entropy used for categorization in CNN models talks about the mathematics applied in the cross-entropy loss and various methods used to apply it. KL divergence, Prediction power and unified loss is discussed in detail.

C.C. Jay Kuo's et al. [10] article on CNN model also gives better understanding of the convolution layer and how this model is purely mathematics. Rectified-correlations on a sphere (RECOS) is mentioned in detail that helps to us to understand the advantage of two-layer cascade system over the on-layer.

Michael K. Buckland et al. [11] published a read on Emanuel Goldberg's work on Electronic Document Retrieval by scanning characters and turning them into telegraph code in 1914. In the late 1920s Emanuel Goldberg developed a "Statistical Machine" using an OCR system.

III. PROPOSED WORK

The main objective is to build a model with the help of which user take the snap, upload in the software, user will select its native language, and then software will convert into user's native language. It will be very useful for the user to read the road instruction board, road signs, Information of any product etc. We need to make a CNN model of more accuracy for character recognition in natural scene images. The main objectives that should be achieved in order for better productivity of tool detect maximum text present in the image, proper segmentation of a character in detected text and providing a tool for translating the detected text into user's native language.

IV. METHODOLOGY

In this section, the methods that are going to be used in this project are described in details.

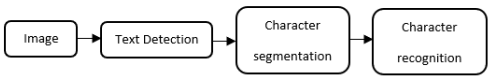


Fig. 1: Flow diagram of the proposed work

The process of character and symbol recognition is generally described in a pipeline as shown in figure 1.

A. Image acquisition

In this process the image containing the text or symbol is acquired by taking a snap of the surrounding.

B. Image pre-processing (processing also includes Text Detection)

The image is then pre-processed by making it ready for operations. It is done by cleaning the image from noise, various filters are added, morphological operations are applied, etc. Then text is detected within the rectangular window.

C. Character Segmentation

In this step the image is segmented to fetch the required characters from it. Segmentation technique is used from digital image processing and individual characters in the image are uniformly resized into 30x20 pixels to make the processing more efficient.

D. Character recognition

This is the decision-making part of the pipeline where neural networks are trained and applied by various techniques like feed forward back propagation, re-enforcement learning algorithms etc.

At last when the characters are recognized the translator detects the language and asks for favorable language to translate to. In this step we would just be using google translator for translation.

Image pre-processing is done in the following steps:

- Image is down-scaled to 320x320 pixel size.
- Gray-scale conversion of the image is done.
- Average threshold method is applied to convert the image into a binary image.
- Delusion and erosion image morphing is done to remove the noises in the image.

E. Pseudo-code used for Character detection and segmentation for further Character recognition

1. *Get the pre-processed image (morphed down-scaled binary image).*
2. *Detect blobs present in the image with the help of method in openCV library.*
3. *Load efficient and accurate scene text (EAST) detector model.*
4. *Loop though the list of blobs and find their respective prediction score and rectangular dimension in the image.*
5. *Ignore the prediction score less than 0.5 for reasonable text detection.*
6. *For each text prediction crop the image out for character segmentation.*
7. *Loop through each cropped image.*
8. *Take a y-strip of an image and find its average.*
9. *If average is not 0 or 1 after an average of 0 or 1 count it as a character encounter and store the x value.*
10. *If average is 0 or 1 the count it as a non-character encounters and stores the value of x.*
11. *Else until a change in average in encountered form the previous average value ignore it.*
12. *Get the list of two groups of x-values in each text image and crop it out.*
13. *Recognize the cropped characters by using a trained CNN model for character detection.*
14. *Combine the characters after looping to identify the text.*

The CNN model used for character detection is made using keras library in python with layers present in the following order:

- Convolution layer-1 with 64 feature detectors of 3x3 pixel size and input image of 32x32 pixel size. Relu activation function is used in this layer.
- Max-Pooling layer of pool size 2x2.
- Convolution layer-2 same as layer 1.
- Flattening layer to convert the feature maps to vectors.
- Full connection layer with 3 ANN layers with relu activation function in hidden layer and softmax function in the output layer.

CNN compiling is done using adam optimizer and categorical cross-entropy loss function. After the text are predicted be use google translate library for language conversion depending on the choices of the user. Google translate switches to a neural machine translation engine – Google Neural Machine Translation (GNMT) to translate the whole sentence or text give as an input. The flowchart of the model is as shown in figure2.

V. RESULTS

The result of the proposed model will be in the form of an application which will ask the user to upload the image to be translated as shown in figure 3.

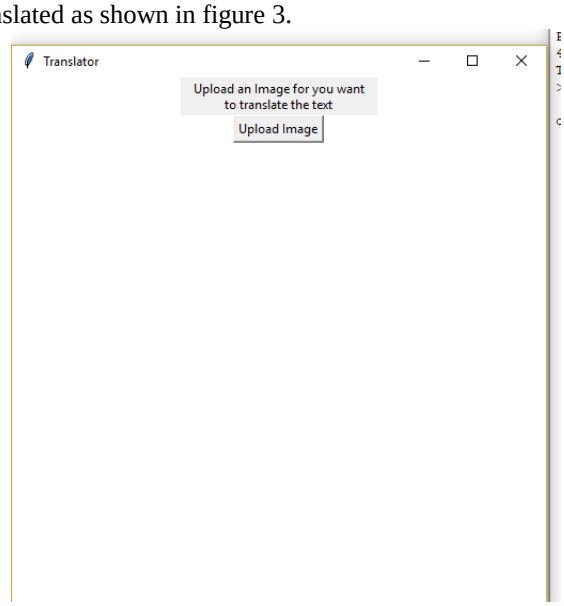


Fig. 3: Upload the image to be translated

After the image will be uploaded it will ask the user to choose the language in which the translation is to be done as shown in figure 4.

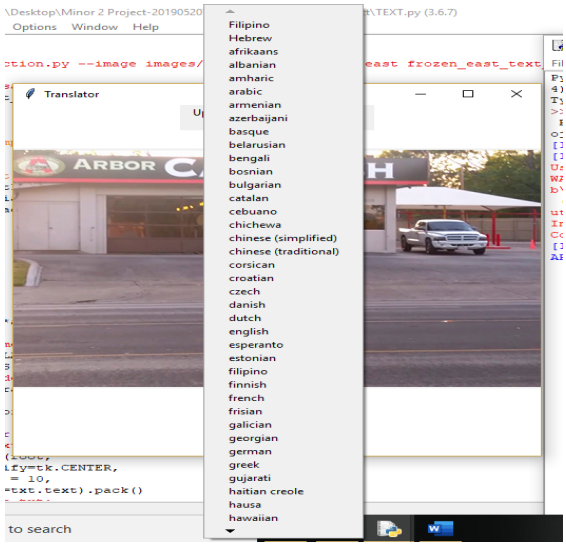


Fig. 4: Language to be selected for translation

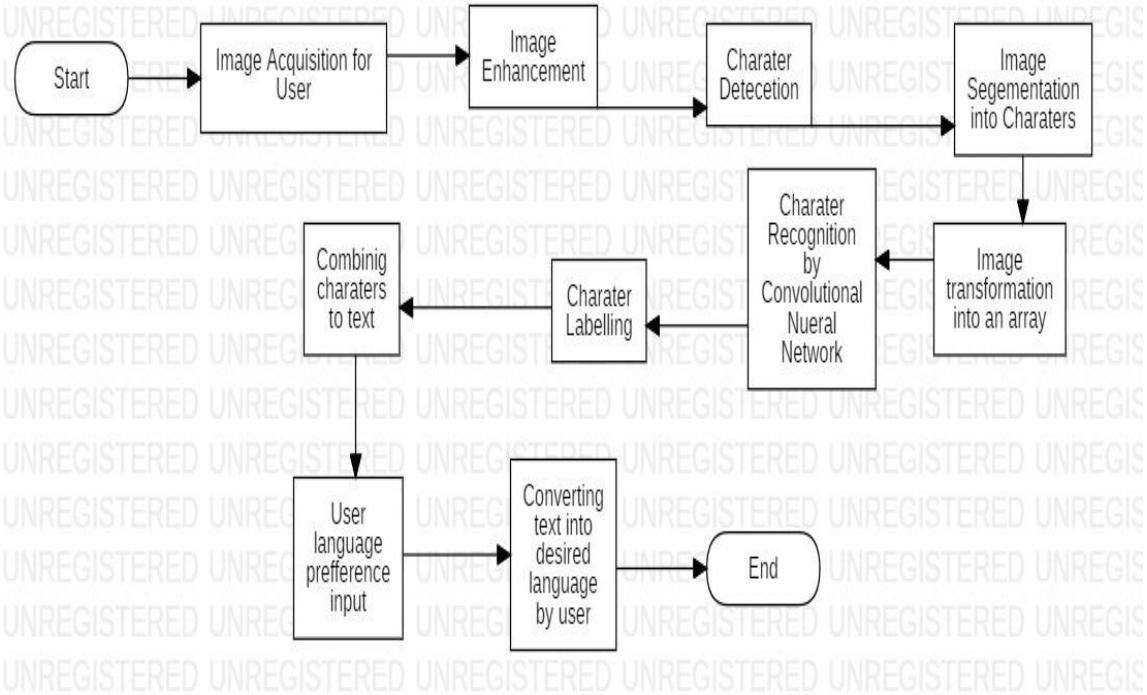


Fig. 2: Flow chart of the model

VI. CONCLUSION

The proposed model will be used to detect the text from natural scene images, extract the text and then convert into user's native language. The model for text segmentation provides segmentation and recognition for natural scene text. It is better due to EAST detector for better text detection, the algorithm is suitable for surrounding text detection as most of them are labels which have spaces between each character, high accuracy of our CNN model which recognize the character.

The future advancements to our work could be enormous. The advancements could be, this proposed algorithm is not suitable for hand writing, small text and closely linked characters detection. Improvement can be done to detect every type of written text. Morphing of an image for noise removal can be improved further in the future for better detection. Accuracy of CNN model character recognition can be increased more. Whole text recognition can be done reduce the separate steps of character segmentation and recognition. Which would decrease the complexity of the algorithm to O (n) where n is the number of text detected in the image.

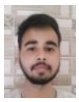
REFERENCES

1. Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). EAST: An Efficient and Accurate Scene Text Detector. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–1, 2017.
2. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET). Doi:10.1109/icengtechnol.2017.8308186
3. Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging, 35(5), 1298.doi:10.1109/tmi.2016.2528162
4. Lin, Tsung-Yu, et al. "Bilinear cnn models for fine-grained visual recognition." Proceedings of the IEEE international conference on computer vision. 2015.
5. Sun, Baohua, et al. "SuperTML: Two-Dimensional Word Embedding and Transfer Learning Using ImageNet Pretrained CNN Models for the Classifications on Tabular Data." arXiv preprint arXiv:1903.06246 (2019).
6. Dürichen, Robert, et al. "Binary Input Layer: Training of CNN models with binary input data." arXiv preprint arXiv:1812.03410 (2018).
7. Yann LeCun et al.,1998, "Gradient-Based Learning Applied to Document Recognition."
8. Jianxin Wu, 2017, "Introction to Convolutional Neural Networks."
9. Rob DiPiertro, 2016, s "A Friendly Inroduction to Cross-Entropy Loss."
10. C.-C. Jay Kuo, 2016, "Understanding Convolutional Neural Networks with A Mathematical Model."
11. Michael K. Buckland, 1992, "Emanuel Goldberg, Electronic Document Retrieval, And Vannervar Bush's Memex." Translated by Banca Fiagicm January 2019.
12. Kakde A. , Arora N., Sharma D, 2019, "A COMPARATIVE STUDY OF DIFFERENT TYPES OF CNN AND HIGHWAY CNN TECHNIQUES" Global Journal of Engineering Science and Research Management 6(4):18-31 DOI: 10.5281/zenodo.2639265

AUTHORS PROFILE

Rishabh Jain is pursuing B. Tech. from University of Petroleum & Energy Studies, Dehradun. His area of interest is cloud computing and Big data.

Saytam Gupta is pursuing B. Tech. Computer Science & Engineering from University of Petroleum & Energy Studies, Dehradun. His area of interest is cloud computing.



Shahzeb Rizvi is pursuing B. Tech. Computer Science & Engineering from University of Petroleum & Energy Studies, Dehradun. His area of interest is cloud computing

Nitin Arora working as an Assistant Professor (SS) with University of Petroleum & Energy Studies, Dehradun. His area of research work includes Image Processing, CBIR, Machine learning. He has published many research papers in National and International Journals.