A Complex Text Detection from Image and Video Frames using Convolutional Neural Network

[1]Gautami D. Dokhe, [2]A.S.Vaidya
[1]Researchscholar(ME- Computer), [2]Assistant Professor
[1] Department of Computer Engineering Gokhale Education Society's,
[1] R. H. Sapat College of Engineering Management Studies and Research,Nashik-5, India

Abstract : With quick heightening of existing multimedia documents and mounting demand for information indexing and retrieval, much endeavor has been done on extracting the text from images and videos. The main objective of the proposition is to spot and get out the scene text from video of different unknown regional languages. These languages are recognized and later translated to English. Extracting the scene text from video is requesting due to complex background, varying font size, different style, lower resolution and blurring, position, viewing angle and so on. Text Region Indicator (TRI) is being developed to compute the text usual confidence and candidate region by performing binarization. Convolutional Neural Networks (CNNs) generally utilized as a part of pattern- and image-recognition problems as they have various favorable circumstances contrasted with different strategies. CNN is a special type of feed-forward multilayer trained in supervised mode. Neural network is used as the classifier and Optical Character Recognition (OCR) is used for character verification. Convolution Neural Network Technique is used for processing. CNN is a class of deep, feed- forward artificial neural networks that has successfully been applied to analyze visual imagery.

IndexTerms **- CNN, Image processing,OCR, Video-to-frames conversion.**

I. INTRODUCTION

Content, a conceptual introduction of manufactured data, is scattered all through the human culture in this compact exhibiting age. As convenient advanced chronicle gadgets are quickly in form among conventional individuals, normal pictures and recordings substance multiply in picture and video sharing sites, e.g. YouTube and Flickr. By removing content data, which conveys abnormal state semantics, characteristic media substance can be adequately comprehended and utilized. It is significant for a wide range of utilizations, for example, picture grouping, scene acknowledgement and programmed route in urban environments. While content acknowledgement in filtered reports has been well studied and effectively sent in genuine applications, the location and acknowledgement of writings in uncontrolled conditions still remains an open issue. By and large, content data extraction can be separated into two phases: content identification and content acknowledgement[1].

II. REVIEW OF LITERATURE

Xiaojun Li et al.[2] proposed a quick and successful way to deal with find content lines even under complex background. Text in pictures and recordings is a huge sign for visual content comprehension and retrieval. First, the calculation utilizes the stroke channel to compute the stroke maps in horizontal, vertical, left-corner to corner, right- slanting headings. At that point a 24-dimensional element is separated for each sliding window and a SVM is utilized to get unpleasant content districts.

Boris Epshtein et al.[3] proposed a novel picture administrator that tries to discover the value of stroke width for each picture pixel, and exhibit its utilization on the undertaking of content identification in normal images. The recommended administrator is nearby and information subordinate, which makes it quick and sufficiently vigorous to wipe out the requirement for multi-scale calculation or examining windows. Broad testing demonstrates that the proposed plot outflanks the most recent published calculations. Its straightforwardness enables the calculation to distinguish messages in numerous text styles and language.

Qixiang Ye et al.[4] proposed framework utilizing multi scale wavelet highlights, proposed a novel coarse-to-fine calculation that can find content lines even under complex background. First, in the coarse identification, after the wavelet vitality highlight is ascertained to find all conceivable content pixels, a thickness based locale developing strategy is created to interface these pixels into locales which are additionally isolated into hopeful content lines by auxiliary data. Also, in the fine identification, with four sorts of surface highlights removed to speak to the surface example of a content line, a forward pursuit calculation is Connected to choose the best highlights.

Weilin Huang et al.[5] proposed a system to deal with this issue by lever maturing the capacity of convolutional neural network (CNN). As opposed to later techniques using a course of action of low-level heuristic highlights, the CNN organize is fit for abnormal state highlights to powerfully distinguish content segments from content like exceptions (e.g.bicycles,windows).
The framework accomplished solid vigor against various extraordinary content varieties and genuine certifiable issues.

Liang Wu et al.[6] proposed another procedure for recognizing and following video writings of any introduction by utilizing spatial and worldly data, respectively. The system explores slope directional symmetry at part level for smoothing edge segments before content discovery. Spatial data is saved by shaping Delaunay triangulation novel at this level, which brings about content hopefuls. Content qualities are then proposed contrastingly to eliminate false content competitors, which comes about in potential content applicants. At that point gathering is proposed for joining potential content.

Cunzhao Shi et al.[7] proposed framework a novel scene content acknowledgement strategy utilizing part-based tree-organized character recognition. Not quite the same as customary multi-scale sliding window character recognition methodology, which does not make utilization of the character particular structure data, we utilize part-based tree-structure to demonstrate each sort of character in order to identify and perceive the characters at the same time. While for word acknowledgement, we manufacture a Contingent Arbitrary Field show on the potential character locations to fuse the identification scores, spatial requirements and etymological information into one structure.

Mehmet Serdar Guzel et al.[8] proposed framework tends to another content acknowledgement arrangement, which is chiefly utilized for the recognition of road see pictures. The utilizes two diverse ways to deal with identify content based locales and perceive relating content fields. The principal approach uses maximally stable extremal districts (MSER), while the second approach depends on the class particular extremal locales (CSER) algorithm. Two isolate systems, planned regarding the previously mentioned techniques, are connected to the road see pictures in order to extricate content based districts. Various examinations were performed to assess and look at both methodologies. Results acquired from the CSER based approach are particularly very promising and check the framework's capacity to distinguish content based areas and perceive comparing content fields.

Tatiana Novikova et al.[9] proposed another model for the assignment of word acknowledgement in common pictures that at the same time models visual and dictionary consistency of words in a solitary probabilistic model. Approach consolidates nearby probability and pair wise positional consistency priors with higher request priors that uphold consistency of characters (lexicon) and their traits (textual style and shading)

Cong Yao et al.[10]proposed Abnormal state semantics exemplified in scene writings are both rich and clear and along these lines can fill in as essential signals for an extensive variety of vision applications, for example, picture understanding, picture ordering, video pursuit, and programmed route. Show a brought together structure for content location and acknowledgement in normal images. As an extra commitment, a novel picture database with writings of distinctive scales, hues, textual styles and introductions in various certifiable scenarios, is created also, discharged. Broad analyses on standard benchmarks and additionally the proposed database show that the proposed framework accomplishes profoundly focused execution, particularly on multi-arranged writings.

III. System Architecture / System Overview

A. Problem Statement

Given an image with complex text , recognize character in it and show text lines as a output using CNN.

B. System Architecture

Fig. 1 shows the architecture of the character recognition System the first step is the image preprocessing stage. The pre-processing is a movement of activities performed on the scanned input image. It fundamentally overhauls the image, making it suitable for further processing. This is an imperative stage as accuracy of the training depends on the quality of processed image. The main advantage of preprocessing a character image is to organize the information so as to make the task of recognition simpler. The next stage is resizing and feature extraction. Finally training of neural network will take place.

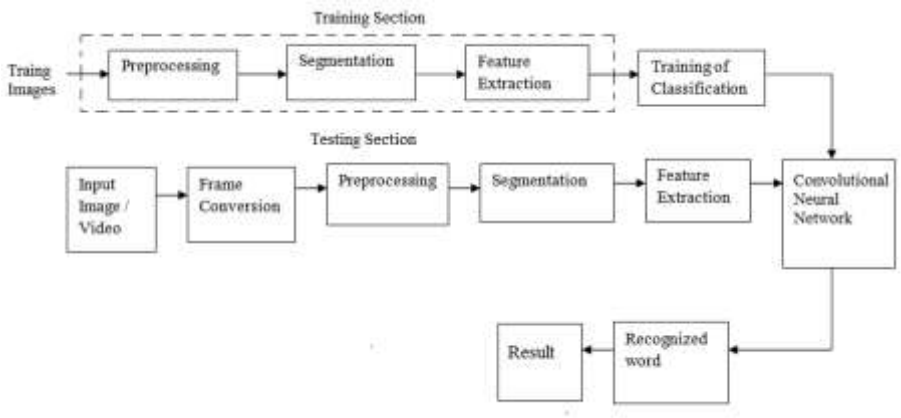

Fig. 1. System Architecture

IV. SYSTEM ANALYSIS

A. Mathematical Model

Let S be a Text detection system, such that,

$S = T, test, F, O, D, T emp, |s$

Where,

T represents set of Training Image, test represents set of Test Image,
F represents Features, O represents OCR,
D represents set of Dataset, Temp represents set of Template

Initial State(S0)

User browse the dataset for Text detection and Recognition

End State(S7)

User obtained the results.

Input

Dataset(D)=d0,d1, .,dn

Output

The relevant results in the text recognized.

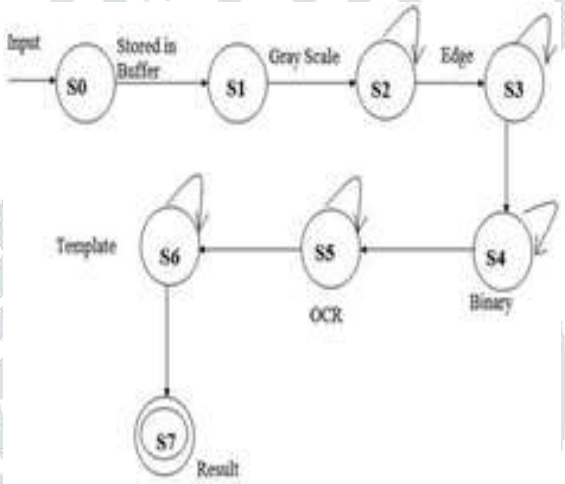

Fig. 2. Mathematical Model

B. Implementation Details Hardware Requirements

There is the new functionality will run on all standards hardware platform like Intel and Macintosh. These frameworks comprise of standard and updated Windows, Apple, and Macintosh working frameworks. Hardware interfaces incorporate ideal for PC with P4 and AMD 64 processor. The base design is required for proposed framework 2.4 GHZ,80 GB HDD for installation and 512 MB memory.

Software Requirements

There are the different service providers will have different software interfaces to access the authentication services provided by the system. They can perform their services independently as long as they adhere with the policies and standard agreed upon. The proposed system uses the software for implementation as JDK 1.7.

V. RESULTS AND IMPLEMENTATION

Fig. 3. Input Image

Here,Accepting the input image from Dataset.

Fig. 4.Image Segmentation Process

Fig shows the preprocessing is a series of operations performed on the scanned input image.

Fig. 5. Blur Background Image.

Fig .Now, further step we are implementing the noise reduction and time reducing character recognition of image. We have to Blur background image and filtering the text.

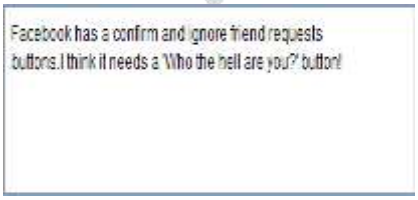

Fig . 6. Result

Fig After applying Optical character recognition algorithm we have to obtain final result in the form of text

A. Performance Measures

We will use the text detection and recognition method for extracted character number(Ext.), character recognition rate(CRR),precision(Prec.) and word recognition rate(WRR) and recognition have to be performed in each frames. In the existing system their does not work on complex character in image but in this system we will work on complex character.

B. Criterions for evaluating the recognition performance

To assess the performance of the different algorithms, we used the character recognition rate (CRR) and the character

precision rate (CPR). They are computed on a ground truth basis as:

$$CRR = \frac{Nr}{N} \quad (1)$$

$$And\ CPR = \frac{Nr}{Ne} \quad (2)$$

where N is the genuine aggregate number of characters, Nr is the quantity of accurately recognized characters and Ne is the aggregate number of removed characters. The quantity of effectively recognized characters is registered utilizing an alter remove between the perceived string and the ground truth.

All the more exactly, let l_T text string, the quantity of cancellations, additions, and substitutions gained when registering the alter separate. The number Nr of accurately perceived characters in this string is then characterized as:

$$Nr = l_T - (del + sub) \quad (3)$$

Naturally, if keeping in mind the end goal to coordinate the ground truth, we have to erase a character or substitute a character, it infers that this character isn't in the ground truth. Additionally, we compute the word recognition rate (WRR) (WRR) to get a thought of the coherency of character acknowledgment inside one arrangement. For every content picture, we check the words beginning starting from the earliest stage of that photo that appear in the string result. In this way,the word recognition rate is defined as the percentage of words from the ground truth that are recognized in the string results.

VI. CONCLUSION

In this system, a novel content structure highlight extractor for both content discovery and acknowledgement utilizing Covolutional Neural System. It well emulates the key instruments in the three-layer content insight model of human, which empowers human distinguish and perceive messages in the meantime. The identification of billboards from road see pictures is a testing errand and requires getting a locale of intrigue (return for capital contributed), including writings and characters. Properly, two distinct models were composed, in view of two diverse division calculations.

ACKNOWLEDGEMENT

I have a tremendous pleasure in presenting the project “A Text Detection and Recognition in Image and video frames using convolutional Neural Network ” under the guidance of Prof. A. S. Vaidya PG coordinator. I am really obligated and appreciative to Head of the Department Dr. D. V. Patil for their significant direction and consolation. I might likewise want to thank the Gokhale Education Society’s

R. H. Sapat College Of Engineering, Management Studies Research, Nashik-5 for giving the required offices, Web get to and vital books. At last I must express my sincere heartfelt gratitude to all the Teaching Non-teachingStaff members of Computer Department of GESRHSCOE who helped me for their important time, support, remarks, thoughts.

REFERENCES

[1] X. Chen and A. L. Yuille, “Detecting and reading text in natural scenes,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2. Jul. 2004, pp. II-366-II-373.

[2] X. Li, W. Wang, S. Jiang, Q. Huang, and W. Gao, “Fast and effective text detection,” in Proc. 15th IEEE Int. Conf. Image Process. (ICIP), Oct. 2008,pp. 969-972.

[3] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.CVPR), Jun. 2010, pp. 2963-2970.

[4] Q. Ye, Q. Huang, W. Gao, and D. Zhao, “Fast and robust text detection in images and video frames,” Image Vis. Comput., vol. 23, no. 6, pp. 565-576,Jun. 2005.

[5] X. Ren, K. Chen, X. Yang, Y. Zhou, J. He, and J. Sun, “A new unsupervised convolutional neural network model for chinese scene text detection,”in Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP),Jul. 2015, pp. 428-432.

[6] W. Huang, Y. Qiao, and X. Tang, “Robust scene text detection with convo- lution neural network induced MSER trees,” in Proc. Eur. Conf.Comput. Vis.(ECCV), 2014, pp. 497-511.

[7] Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang, Song Gao and Zhong Zhang, “Scene Text Recognition using Part-based Tree-structured Character Detection,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition PP. 2961-2968, Jun. 2013.

[8] Mehmet Serdar Guzel, “A Novel Framework for Text Recognition in Street View Images,” in Proc. International Journal of Intelligent Systems and Applications in Engineering, Vol 5, No 3, Sep. 2017.

[9] Tatiana Novikova, Olga Barinova, Pushmeet Kohli, Victor Lempitsky, "Large-Lexicon Attribute-Consistent Text Recognition in Natural Im- ages," European Conference on Computer Vision ECCV 2012: Computer Vision ECCV 2012, pp. 752-765.

[10] Cong Yao, Xiang Bai, Member, IEEE, and Wenyu Liu, Member, IEEE, "A Unified Framework for Multi-Oriented Text Detection and Recognition," in Proc. IEEE Transactions on Image Processing, Volume: 23, Issue: 11, Nov. 2014 , pp. 4737 - 4749

[11] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, "Reading Text in the Wild with Convolutional Neural Networks," in Pro. Computer Vision and Pattern Recognition, 4 Dec. 2014.

[12] Yingying ZhuCong YaoXiang Bai, "Scene text detection and recogni- tion: recent advances and future trends," Frontiers of Computer Science February2016, Volume 10, Issue 1, pp 1936.

[13] Huizhong Chen, Sam S. Tsai, Georg Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod, "robust text detection in natural images with edge-enhanced maximally stable extremal regions," in Proc. Image Processing (ICIP), 2011 18th IEEE International Conference on Sept. 2011

[14] [14] Hojin Cho, Myungchul Sung, Bongjin Jun, "Canny Text Detector: Fast and Robust Scene Text Localization Algorithm," in Proc. Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on 27-30 June 2016.

[15] Cong Yao, Xin Zhang, Xiang Bai, Member, IEEE, Wenyu Liu, Member, IEEE, Yi Ma, Senior Member, IEEE, and Zhuowen Tu, Member, IEEE, "Detecting Texts of Arbitrary Orientations in Natural Images," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1083-1090.

[16] Tong He, Weilin Huang, Member, IEEE, Yu Qiao, Senior Member, IEEE, and Jian Yao, Senior Member, IEEE , "Text-Attentional Convo- lutional Neural Network for Scene Text Detection," in Proc. Computer Vision and Pattern Recognition, Oct. 2015.

[17] L. Wu, P. Shivakumara, T. Lu, and C. L. Tan, "A new technique for multi-oriented scene text line detection and tracking in video," IEEE Trans.Multimedia, vol. 17, no. 8, pp. 1137-1152, Jan. 2015.

[18] A. Mishra, K. Alahari, and C. Jawahar, "Scene text recognition us- ing higher order language priors," in Proc. Brit. Mach. Vis. Conf. (BMVC),2012, p. 1.

[19] C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multi-scale representation for scene text recognition," in Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR), Jun. 2014, pp. 4042-4049.

[20] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet,(2013). "Multidigit number recognition from street view im- agery using deep convolutional neural networks," [Online]. Available: https://arxiv.org/abs/1312.6082.

[21] C. Shi, C.Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, "Scene text recognition using part-based tree-structured character detection,"in Proc. IEEE Conf.Comput. Vis. Pattern Recognit. (CVPR), Apr. 2013,pp. 2961- 2968.

[22] T. Novikova, O. Barinova, P. Kohli, and V. Lempitsky, "Large-lexi conattribute-consistent text recognition in natural images," in Proc. Eur. Conf.Comput. Vis. (ECCV), 2012, pp. 752-765.

[23] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recog- nition,"in Proc. Comput. Vis. (ICCV), 2011, pp. 1457-1464.

[24] L. Neumann and J.Matas, "On combining multiple segmentations in scene text recognition," in Proc. 12th Int. Conf. Document Anal. Recognit. (ICDAR),Oct. 2013, pp. 523-527.

[25] C. Yao, X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," IEEE Trans. Image Process., vol. 23, no. 11,pp.4737-4749, Jun. 2014.

[26] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," Int. J. Comput. Vis.,vol. 116,no.1,pp.1-20,2016.