International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I1P220
Volume 70 Issue 1, 180-184, January, 2022

Detection of Text from Lecture Video Images using CTPN Algorithm

Geeta S Hukkeri[1], R H Goudar[2]

[1]Research Scholar, Department of Computer science and Engineering, Visvesvaraya Technological University, Belagavi, India.

[2] Associate Professor, Department of Computer science and Engineering, Visvesvaraya Technological University, Belagavi, India

[1]geetahukkeri7@gmail.com, [2]rhgoudar.vtu@gmail.com

Abstract — *Text data in lecture video images play a crucial indication in understanding lecture videos. Different stages are involved in the detection of such data from images that are pictured by text detection methods. The goal of this paper is to study the existing text detection methods and reveal the best method among them. This paper presents a deep learning-based methodology known as CTPN that precisely confines text lines in a common picture in the content discovery stage. The CTPN operates dependably on multi-scale and multi-language text minus any additional post-preparing, withdrawing from past base-up techniques needing multi-step post filtering. The result of detected text has been shown in this paper.*

Keywords — *Text detection, text localization, Conventional approach, CTPN Algorithm.*

I. INTRODUCTION

Different applications in the field of picture handling and PC vision need to extricate the substance in a given picture. Content in a picture is basically isolated into two primary classes: a) semantic substance and b) perceptual substance. Semantic substance adds to different articles, occasions, and their relations. It manages the connection between words, expressions, signs, and images. In a basic way, semantic alludes to the investigation of significance, whereas perceptual substances are a substance that can be deciphered; such substance incorporates properties like tone, force, shape, surface, and their fleeting changes. Heaps of work have been accounted for on perceptual substance. Semantic picture content as text, face, vehicle, and human activity are drawn in some new interest. In this paper, the target has been given around semantic picture content as text since text inside a picture is specifically compelling as it is extremely helpful for portraying the substance of a picture.

At the point when OCR is applied to regular scene pictures, the achievement pace of OCR drops radically because of a few reasons [15]. To begin with, most OCR tools are intended for checked content, which utilizes division to isolate out text from background pixels. Second, a unique assortment of text designs like sizes, text styles, directions, colors, presence of background anomalies like content characters, like windows, blocks, and character-like surfaces. At long last, page format for conventional OCR is basic and organized; however, in regular scene pictures, there is undeniably less content, and there exists generally structure with high variation both in math and appearance. In light of these issues, scene text pictures, for the most part, experience the ill effects of photometric corruptions just as mathematical mutilation, with the goal that numerous calculations confronted the exactness or potentially speed issue. Hence, the main errand while planning applications utilizing text data in regular scene pictures is to change the picture such that current OCR tools can oversee.

The remnant of the paper is coordinated as follows; section 2 gives brief related work on text detection modules. Section 3 gives an outline of the application space for TD in picture handling and a comparative study on different text detection methods. Section 4 portrays CTPN Algorithm with implementation, and 5 gives the conclusion.

II. RELATED WORK

Motivated by the great spectrum of latent applications, research topics on scene text detection and recognition have become very important [24], [25], [36]–[42] and have newly observed a rise in research attempts in the computer vision group [26]–[31], [32], [33], [34], [35], [43]–[45]. Broad studies on text detection and recognition in still images and videos can be discovered in [46]–[48].

Bottom-up methods have been used for text detection in earlier works. “Connected-components (CCs)” based methodologies and “sliding-window” based strategies are the two types typically collected. The CCs based methodologies set apart text and non-text picture factors by utilizing a quick strain. Text factors are put together into a character field by utilizing low-level features, e.g., force, shading, angle, and so forth [17, 21, and 22].

The sliding–window–based approaches recognize text fields by closely stimulating a multi-scale window over a picture. The text or non-text window is separated by a pre-designed classifier by applying manually-designed attributes [19, 20] or previous CNN highlights [18]. Different kinds of text indicators have been considered. Lately, that may be widely classified into two styles, namely, “normal text detectors” and “deep learning-based text detectors”. Normal content indicators influence unreal minor features and older information to identify text and non-text portions in the picture. Though, these methods lack the strength of different text writings and unclean pictures. To alleviate such an issue, much deep learning-based content identification investigations have been done, and they could accomplish best [14].

III. TEXT DETECTION (TD) SYSTEM

Text in pictures and video can be ordered into two classes: scene text and caption text. Figure 1 shows the instances of scene text and caption text. Caption text is unnaturally covered on the pictures or video outlines while editing. Even if a few kinds of caption text are more normal than others, it is seen that caption text can have self-assertive textual style, size, shading, direction, and area. For video record, a caption text occasion might stay fixed and inflexible after some time, or it might interpret, develop, and change the tone.

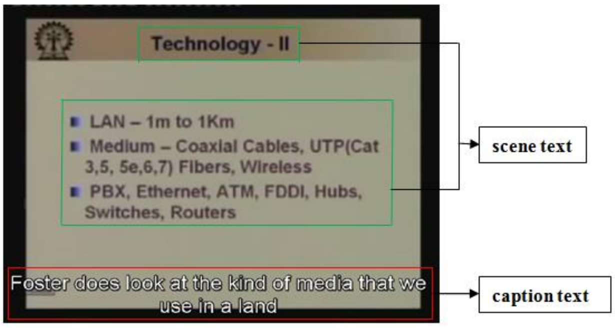


Fig.1 Instances of caption text and scene text.

A good text detection framework should have the option to deal with as wide an assortment of these sorts of text as conceivable to help a content-based indexing framework. So it is advisable to apply a general-purpose text detection system (TD), shown in Figure 2.

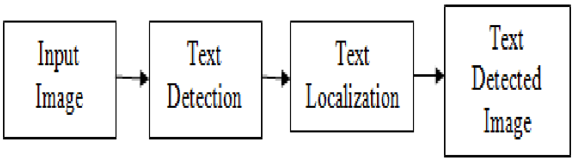


Fig.2 TD System

- An input image is sent to the TD system. The TD system has two parts to it, namely, text detection and text localization.
- *The text detection* part checks whether the text exists in a given image or not.
- *The text localization* part determines the location coordinates of the resided text in an image and creates bounding boxes over the text.

Once the text is detected from an image that can be extracted in the form of normal text, apply a good OCR method.

Table.1 Text detection methods

Method	Advantage	Disadvantage
SWT[5][7][8] [11]	Correctly detect text in various colors, sizes, styles. Text detection can be done independently of position and rotation	Slow performance. If background or text is darker, information is needed (in the grayscale)
DTLN [1][12]	It recurrently memorizes the previous detection to avoid repeated detection of text at multiple scales and efficiently catches crowded text samples in close proximity. Works robustly in noisy environments	Slow performance
EAST [2][13]	Written in Keras, which helps to read and run easily.	It can detect a single word
MSER [5][6][8] [11]	Correctly detect text in various colors, sizes, styles. Text detection can be done independently of position and rotation. Performance is good.	Sensible in comparison with blur.
CTPN [3][4][9] [10][14]	Correctly localizes text lines in an image. Works accurately on multi-language and multi-scale text using no further post-processing. Best performance due to the use of bidirectional LSTM and text line construction algorithm	Tilted text boxes are not detected accurately like EAST.

IV. CTPN ALGORITHM

This paper presents a new Connectionist Text Proposal Network (CTPN) this quickly localizes text alliances in convolutional layers [14]. This solves various primary impediments built by earlier bottom-up methods making on text detection. This paper pulls the favors of solid deep convolutional highlights and presents the CTPN framework shown in Fig. 3. It forms the subsequent significant commitments: Firstly, from the point of text discovery into localizing a series of "fine-scale text proposals" [16]. This paper progresses an anchor regression system that mutually forecasts the text/non-text count of every content proposition, bringing about amazing localization correctness. Secondly, the paper presents an "in-network recurrence" method that richly associates successive content recommendations in the convolutional feature maps. This alliance lets the detecting device investigate significant text line data, making it incredible to recognize text reliably. Thirdly, the two techniques are coordinated consistently to satisfy the concept of text series, causing a cooperative end-to-end compliant model. This strategy can deal with multi-scale and multi-spoken content in a solitary cycle, keeping away from additional post filtering or refinement.

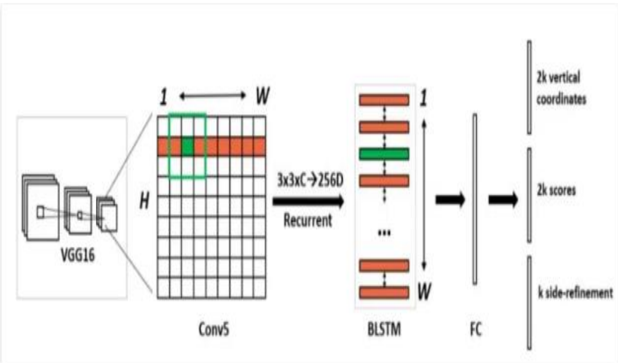


Fig. 3 Framework of the (CTPN)

The input image is being passed to the VGG-16 model [23]. Features yield from conv_5 layer of VGG-16 model is obtained. A sliding window of size 3X3 is moved over VGG-16 yield features and afterward taken care of consecutively to the RNN network, which comprises 256D bi-directional LSTM. This LSTM layer is associated with a 512D fully connected layer, which will then generate the output layer.

A. Implementation

This paper implements the CTPN algorithm using its GitHub repository to localize text areas in an image. We have used Ubuntu operating system for this work. The steps are as follows:

a) Clone the Repository: Clone the CTPN GitHub Repo using the following command:

```
git clone https://github.com/eragonruan/text-detection-ctpn.git
```

b) Build the Required Library: Generate .so file for NMS (Non-max suppression) and bbox (bounding box) utilities so that vital files can be loaded into the library.

The existing directory has to be changed to "/text-detection-ctpn/utils/bbox" using the following command:

```
cd text-detection-ctpn/utils/bbox
```

Now build the library using the following commands.

```
chmod +x make.sh
./make.sh
```

c) Test the model

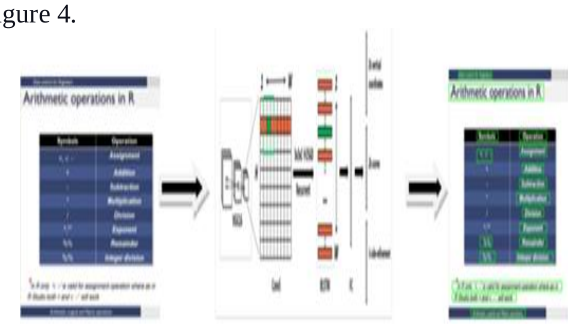
Test the CTPN model by downloading the checkpoints (available in the GitHub repository). Use the below steps:

1) Open the downloaded checkpoints.
2) Locate the opened folder "checkpoints_mlt" in "/text-detection-ctpn" directory.
3) Place testing images in the "/data/demo/" folder, and results will be produced in the "/data/res" folder.

Change directory to "/text-detection-ctpn" and run the below command to test the input images.

```
cd /text-detection-ctpn
python main/demo.py
```

Results have been produced on the "/data/res" folder. The sample input image and its result are shown in below figure 4.



(a) Input **(b) Output**

Fig.4 Text detection from lecture video image using CTPN

V. CONCLUSION

This paper talk about two types of images, caption text image and scene text image. This paper dealt with the scene text image and presented a general-purpose text detection (TD) system, which performs detection and localization of the text in a given image. In this paper, a study on different text detection methods is discussed. From the study of several papers, we found that the CTPN algorithm gives more accurate text detection compared to other methods. The result acquired of text detection is truly exact and very usefull.

ACKNOWLEDGMENT

The satisfaction and happiness that accompany the successful completion of any task would be complete without the mention of people who made it possible, whose constant guidance and encouragement crowned our effort with success.

We are very thankful to **Dr. Karisiddappa**, Vice-Chancellor, VTU Belagavi, who has supported, motivated, and encouraged us to write this paper.

We are indeed grateful to **Dr. S L Deshpande,** Chairperson, VTU-Belagavi, who has guided and provided all the necessary guidance and support to complete this work.

We would like to express our special thanks to **Mr. Rohit Kaliwal** and **Dr. Mahantesh Birje,** Centre for Post-Graduate Studies, VTU-Belagavi, for their guidance, support, motivation, and encouragement to write this paper.

Furthermore, We acknowledge the support of our parents, family, and friends

REFERENCES

[1] Nils L. Westhausen1 and Bernd T. Meyer, Dual-Signal Transformation LSTM Network for Real-Time Noise Suppression, SPRINGER, 1-5 (2020).

[2] Niddal h. Imam, Vassilios g. Vassilakis, An approach for detecting image spam in OSNs, SPRINGER, (2020) 1-8.

[3] Langcai Cao, Hongwei Li, Rongbiao Xie, And Jinrong Zhu, A Text Detection Algorithm for Image of Student Exercises Based on CTPN and Enhanced YOLOv3, IEEE Access, 8 (2020) 176924-176934.

[4] Xiangwen Liu, Joe Meehan, Weida Tong, Leihong Wu, Xiaowei Xu, and Joshua Xu, DLI-IT: a deep learning approach to drug label identification through image and text embedding, BMC Medical Informatics and Decision Making, (2020) 1-9.

[5] Anurag Agrahari, Rajib Ghosh, Multi-Oriented text detection in natural scene images based on the intersection of MSER with the locally binarized image, Third International conference on computing and network communications (CoCoNet' 19), Elsevier, (2020) 322–330.

[6] Rituraj Soni, Bijendra Kumar, and Satish Chand, Text Region Extraction From Scene Images Using AGF and MSER, International Journal of Image and Graphics, 20(2) (2020) 1-23.

[7] A. Vishnu Vardhan, Padarthi Maheeja, Manukonda SaiSudhanvi, Nuthalapati Preethi, Mannava Veena, Detection of Text by Enhancing Stroke Width Transform and Maximally Stable Extremal Regions, © 2020 JCRT, 8(3) (2020).

[8] S.Shiyamala, S.Suganya, Detection, Localization of Text in Images by Mser and Enhanced Swt, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8 (2019) 2873-2875.

[9] Wenxian Zeng, Qinglin Meng, Shuqing Zhang, Natural Scene Chinese Character Text Detection Method Based on Improved CTPN, IOP Conf. Series: Journal of Physics: Conf. Serie, (2019) 1-7.

[10] Chenhui Huang, Jinhua Xu, An Anchor-Free Oriented Text Detector with Connectionist Text Proposal Network, Proceedings of Machine Learning Research, (2019) 631.

[11] JIANG Mengdi, CHENG Jianghua, CHEN Minghui, and KU Xishu, An Improved Text Localization Method for Natural Scene Images, IOP Conf. Series: Journal of Physics: Conf. Series, (2018) 1-8.

[12] Xuejian Rong, Chucai Yi, and Yingli Tian, Unambiguous Text Localization and Retrieval for Cluttered Scenes, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017) 5494.

[13] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, J. Liang, EAST: an efficient and accurate scene text detector, in Proceedings of the CVPR, (2017) 2642.

[14] Z. Tian, W. Huang, T. He, P. He, Y. Qiao, Detecting text in natural image with connectionist text proposal network, in Proceedings of the ECCV, (2016) 56.

[15] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform, CVPR, (2010) 2963.

[16] H. Lin, P. Yang, and F. Zhang, Review of scene text detection and recognition, Arch. Comput. Methods Eng., 27(2) (2020) 433–454.

[17] Z. Tian, W. Huang, T. He, P. He, Y. Qiao, Detecting text in natural image with connectionist text proposal network, in Proceedings of the ECCV, (2016) 56.

[18] Jaderberg, M., Vedaldi, A., Zisserman, A., Deep features for text spotting, in European Conference on Computer Vision (ECCV), (2014) 512.

[19] Tian, S., Pan, Y., Huang, C., Lu, S., Yu, K., Tan, C.L., Text flow: A unified text detection system in natural scene images, in IEEE International Conference on Computer Vision (ICCV), (2015) 1.

[20] Wang, K., Babenko, B., Belongie, S., End-to-end scene text recognition, in IEEE International Conference on Computer Vision (ICCV), (2011) 1457.

[21] Huang, W., Lin, Z., Yang, J., Wang, J., Text localization in natural images using stroke feature transform and text covariance descriptors, in IEEE International Conference on Computer Vision (ICCV), (2013) 1241.

[22] Huang, W., Qiao, Y., Tang, X., Robust scene text detection with convolutional neural networks induced mser trees, in European Conference on Computer Vision (ECCV), (2014) 497.

[23] Simonyan, K., Zisserman, A., Very deep convolutional networks for large-scale image recognition, in International Conference on Learning Representation (ICLR), (2015) 1.

[24] D. Chen, J.-M. Odobez, and H. Bourlard, Text detection and recognition in images and video frames, Pattern Recognit., 37(3) (2004) 595–608.

[25] M. R. Lyu, J. Song, and M. Cai, A comprehensive method for multilingual video text detection, localization, and extraction, IEEE Trans. Circuits Syst. Video Technol.,15(2) (2005) 243–255.

[26] T. E. de Campos, B. R. Babu, and M. Varma, Character recognition in natural images, in Proc. VISAPP, (2009) 121–132.

[27] B. Epshtein, E. Ofek, and Y. Wexler, Detecting text in natural scenes with stroke width transform, in Proc. IEEE CVPR, (2010) 2963.

[28] K. Wang and S. Belongie, Word were spotting in the wild, in Proc. 11th ECCV, (2010) 591.

[29] L. Neumann and J. Matas, A method for text localization and recognition in real-world images, in Proc. 10th ACCV, (2010) 770.

[30] C. Yi and Y. Tian, Text string detection from natural scenes by structure-based partition and grouping, IEEE Trans. Image Process., 20(9) (2011) 2594–2605.

[31] A. Coates et al., Text detection and character recognition in scene images with unsupervised feature learning, in Proc. ICDAR, (2011) 440.

[32] Neumann and J. Matas, Real-time scene text localization and recognition, in Proc. IEEE CVPR, (2012) 3538.

[33] T. Novikova, O. Barinova, P. Kohli, and V. Lempitsky, Large-lexicon attribute-consistent text recognition in natural images, in Proc. 12th ECCV, (2012) 752.

[34] A. Mishra, K. Alahari, and C. V. Jawahar, Top-down and bottom-up cues for scene text recognition, in Proc. IEEE CVPR, (2012) 2687.

[35] A. Ikica and P. Peer, An improved edge profile based method for text detection in images of natural scenes, in Proc. IEEE EUROCON, (2011) 1.

[36] R. Lienhart and F. Stuber, Automatic text recognition in digital videos, Dept. Math. Comput. Sci., Univ. Mannheim, Mannheim, Germany, Tech. Rep, (1995).

[37] Y. Zhong, K. Karu, and A. K. Jain, Locating text in complex color images, Pattern Recognit., 28(10) (1995) 1523–1535.

[38] V. Wu, R. Manmatha, and E. M. Riseman, Finding text in images, in Proc. 2nd ACM Int. Conf. Digital Libraries, (1997) 3.

[39] A. K. Jain and B. Yu, Automatic text location in images and video frames, Pattern Recognit., 31(12) (1998) 2055.

[40] V. Wu, R. Manmatha, and E. M. Riseman, Textfinder: An automatic system to detect and recognize text in images, IEEE Trans. Pattern Anal. Mach. Intell., 21(11) (1999) 1224–1229.

[41] Y. Zhong, H. Zhang, and A. K. Jain, Automatic caption localization in compressed video, IEEE Trans. Pattern Anal. Mach. Intell., 220(4) (2000) 385–392.

[42] H. Li, D. Doermann, and O. Kia, Automatic text detection and tracking in digital video, IEEE Trans. Image Process., 9(1) (2000) 147–156.

[43] Y.-F. Pan, X. Hou, and C.-L. Liu, A hybrid approach to detect and localize texts in natural scene images, IEEE Trans. Image Process., 20(3) (2011) 800–813.

[44] C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, and Z. Tu, Rotation-invariant features for multi-oriented text detection in natural images, PLoS ONE, 8(8) (2013) 1-15.

[45] X.-C. Yin, X. Yin, K. Huang, and H.-W. Hao, Accurate and robust text detection: A step-in for text retrieval in natural scene images, in Proc. 36th SIGIR, (2013) 1091.

[46] D. Chen, J. Luettin, and K. Shearer, A survey of text detection and recognition in images and videos, Institut Dalle Molle d' Intelligence Artificielle Perceptual, EPFL, Martigny, Switzerland, Tech. Rep. (2000) 1-21.

[47] K. Jung, K. Kim, and A. Jain, Text information extraction in images and video: A survey, Pattern Recognit., 37(5) (2004) 977–997.

[48] J. Liang, D. Doermann, and H. Li, Camera-based analysis of text and documents: A survey, Int. J. Document Anal. Recognit., 7(2) (2005) 84–104.

[49] M Krishna Goriparthy and B Geetha Lakshmi, Balanced Islanding Detection of Distributed Generator using Discrete Wavelet Transform and Artificial Neural Network, International Journal of Engineering Trends and Technology, 69(10) (2021) 57-71.

[50] Shankargoud Patil and Kappargaon S. Prabhushetty, Bi-Attention LSTM with CNN based Multi-task Human Activity Detection in Video Surveillance, International Journal of Engineering Trends and Technology, 69(11) (2021) 192-204.

[51] Vencelin Gino V and Amit KR Ghosh, Enhancing Cyber Security Measures For Online Learning Platforms, SSRG International Journal of Computer Science and Engineering, 8(11) (2021) 1-5.

[52] Neha Chaudhary and Priti Dimri, Enhancing the Latent Fingerprint Segmentation Accuracy Using Hybrid Techniques – WCO and BiLSTM, International Journal of Engineering Trends and Technology,. 69(11) (2021) 161-169.

[53] Emmanoel Pratama Putra Hastono and Gede Putra Kusuma, Evaluation of Blockchain Model for Educational Certificate Using Continuous-Time Markov Chain, International Journal of Engineering Trends and Technology, 69(11) (2021) 61-70.

[54] Ningxia He, Image Sampling Using Q-Learning, SSRG International Journal of Computer Science and Engineering, 8(1) (2021) 5-12,.

[55] Walther Calvo-Niño, Brian Meneses-Claudio, and Alexi Delgado, Implementation of an FM Radio Station to Facilitate Remote Education in the District of Cojata, International Journal of Engineering Trends and Technology, 69(10) (2021) 118-127.

[56] Osama R.Shahin, Zeinab M. Abdel Azim, and Ahmed I Taloba, Maneuvering Digital Watermarking In Face Recognition , International Journal of Engineering Trends and Technology, 69(11) (2021) 104-112.

[57] Yasser Mohammad Al-Sharo, Amer Tahseen Abu-Jassar, Svitlana Sotnik, and Vyacheslav Lyashenko, Neural Networks As A Tool For Pattern Recognition of Fasteners, International Journal of Engineering Trends and Technology, 69(10) (2021) 151-160.

[58] Hoang Thi Phuong, Researching Robot Arms Control System Based On Computer Vision Application And Artificial Intelligence Technology, SSRG International Journal of Computer Science and Engineering, 8(1) (2021) 24-29.

[59] Nilesh Yadav and Dr. Narendra Shekokar, SQLI Detection Based on LDA Topic Model, International Journal of Engineering Trends and Technology, 69(11) (2021) 47-52.

[60] M. Ayaz Ahmad, Volodymyr Gorokhovatskyi, Iryna Tvoroshenko, Nataliia Vlasenko, and Syed Khalid Mustafa, The Research of Image Classification Methods Based on the Introducing Cluster Representation Parameters for the Structural Description, International Journal of Engineering Trends and Technology, 69(10) (2021) 186-192.