# Study on Text Detection and Recognition from Images

Dr. Dinesh D. Patil
*Associate Professor, CSE, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal*

***Abstract*** **- In this system try and extract words from pictures and take into account character segmentation as a language-independent and individual character recognition as a language-dependent downside. Uses a deep learning approach to extract words from pictures. For text recognition, a concatenation structure is meant to affix the options from each shallow and deep layers in neural networks. Scene text recognizing aims at coinciding localization and recognizing text instances, symbols, and logos in natural scene pictures. text detection and recognition approaches have received immense attention in the computer vision research community.**

***Index Terms*** **- CNN, Detection, Neural network, Textual information extraction, Recognition.**

## I.INTRODUCTION

As autonomous vehicles and other intelligent devices like mobile or Robots are coming online, and they also need to understand the environment in which they are operating. Different kind of instructions for people's guidance is often written everywhere in public places, especially on different signboards, hoardings, banners along the roads. High denser digital cameras used in smartphones capture natural scene images of the world around us in real time. Natural scenes contain both text and various scene objects. Texts embedded in scene images contain a large amount of useful information. Unlike the characters in printed documents, image texts are more difficult to recognize due to the large variations in backgrounds, textures, fonts, and illumination conditions. Multi-language text spotting is an essential but challenging task. This is a challenging task because spotting results can be significantly affected by a wide variation in size, orientation, aspect ratio, color, script, and font of text instances in the images. Text extraction from an image is a technique uses machine learning to extract the text directly from the picture. Thinking of text extraction from images is thinking of a way to teach AI algorithms how to read.

## II.LITRETURE REVIEW

### A. TEXT DETECTION

The early approaches mostly follow a bottom-up pipeline that applies artificial features to detect strokes or characters[13]. The individual character or combined strokes are directly classified in the recognition period or constructed into a line for text line verification in text detection[1].However, their performance relies on the results of character detection, and the extracted features are not robust to distinguish strokes or characters in different scenes (e.g., various fonts and degraded images)[3].

These methods can be categorized into bounding box regression-based methods, segmentation-based methods, and combined methods[2][4]. Bounding box regression-based methods treat each textual area as a kind of object and directly predict its bounding box and classification. Segmentation based methods try to segment text regions from the background and output the final bounding boxes according to the segmented results[6][7]. Combined methods use a similar approach like Mask R-CNN, in which both segmentation and bounding box regression are used for better performance. However, combined methods are time consuming because more steps are involved. Among the three kinds of methods, bounding box regression-based methods are the most popular in scene text detection and also adopt this kind of method[7][10][12].

Bounding box regression-based methods can be divided into one-stage methods and two-stage methods. One-stage methods directly output detection results at several grids that correspond to the specific locations on feature maps[18].

Considering the sequence characteristic of text, CTPN (Connectionist Text Proposal Network) combines CNN and RNN (Recurrent Neural Network) to detect sequential features. EAST (An Efficient and Accurate Scene Text Detector) is another two-stage detector, where a FCN-based (Fully Convolutional Networks )

pipeline is devised to merge the features from each convolutional layer.

B. TEXT RECOGNITION

Traditional methods recognize characters individually and then group them into words. They explore low-level features, which are not robust to identify complex structures without context information. However, it is challenging to segment single characters because of the complicated background and inconsistent character spacing, e.g Chinese character spacing is usually larger than Latin character spacing. The CTC (Connectionist Temporal Classification) loss is connected with the RNN outputs for calculating the conditional probability of the prediction [17] [18].

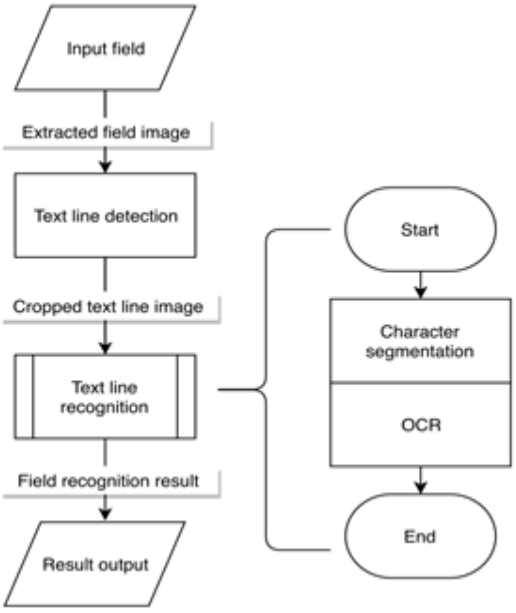


Figure 1 Block diagram Text detection & Reorganisation

III. METHODOLOGY

Due to the matter of bilingual texts and restricted real information, the recognizer is increased with the projected concatenation structure and trained on an artificial dataset. The output for one supply image contains the localizations and contents of all detected texts.

*A.* Text Detection in Document Image

The supervised feature learning and end-to-end training procedure create it straightforward to transfer neural network strategies to alternative applications. Adopt a two-stage design that's originally used for generic object detection. The patch based mostly strategy is applied to the present design for text detection. Multiple optimizing strategies are adopted during this work to enhance the performance.

*B.* Multi-Lingual Text Recognition

Deep convolutional networks can learn high-level features through successive convolutions.

In the network, the options from 2 adjacent convolution layers are going to be concatenated along because the input of the third layer. Every convolution connects with ReLU (Rectified Linear Unit) perform. From the third layer, the input of every convolution layer is that the concatenation of its previous 2 layers' outputs. Average pooling is employed here to squeeze the feature maps so they'll have constant dimension and height before concatenation. it's noted that this operation doesn't bring too several further parameters compared with convolution or de convolution.

2. Training

Given a batch of textual images, they are resized into $(h_0 \times w_0)$,

Where $h_0 = 32$, and $w_0$ is the maximum width among these images.

Then the batch of images is fed into the network, which outputs a sequence of labels

$$\hat{y}^i = \hat{y}^1. \hat{y}^m., \text{Each } \hat{y}^i \in D,$$

where, D is the dictionary that contains all characters in our task.

Because the prediction may include incorrect labels, repeated labels, and 'blank's, and adopt the conditional probability defined in the Connectionist Temporal Classification (CTC) layer to align the prediction and ground truth.

- A patch-based strategy is employed for text detection on documentary pictures with high resolution and this strategy leads to a high recall and exactitude.
- A concatenation structure is projected that mixes the options from 2 adjacent convolution layers and brings a major improvement in a very multilingual scene.
- A deep learning approach is given for text detection and recognition from pictures of medical laboratory reports.

C.Test with Multiple Resolutions

The image resolution is around 2500 × 3400. The model trained on such a dataset could also be over fitting. so as to verify the lustiness of the projected approach, we have a tendency to generate a multi-resolution check set wherever every original check image has 5 completely different resolutions. Every new image within the multiresolution check set is obtained by resizing a resourceful check image with a scale indiscriminately drawn. For text detection, quicker RCNN and EAST square measure chosen as comparison ways for his or her smart performance

## IV. CONCLUSION

Study on this paper shows a deep learning approach for text detection and recognition from pictures. First, a patch-based coaching is applied to a detector that outputs a collection of bounding boxes containing texts. Then a concatenation structure is inserted into a recognizer, that takes the areas of bounding boxes in supply image as inputs and outputs recognized texts. In text detection experiments, image resolution will seriously have an effect on the detection results. The popularity experimental results demonstrate that the concatenation structure will effectively mix shallow and deep options and contribute to the popularity performance. Large size datasets with a lot of transformations covering noise, rotation, color, background, font variations are targeted, needed for deciding the lustiness of these techniques.

## REFERENCES


[1] C. Rossignoli, A. Zardini, and P. Benetollo, ‘‘The process of digitalisation in radiology as a lever for organisational change: The case of the academic integrated hospital of verona,’’ in DSS 2.0—Supporting Decision Making with New Technologies, vol. 261. Amsterdam, The Netherlands: IOS Press, 2014, pp. 24–35.

[2] S. Bonomi, ‘‘The electronic health record: A comparison of some European countries,’’ in Information and Communication Technologies in Organizations and Society. Cham, Switzerland: Springer, Jan. 2016, pp. 33–50.

[3] A. K. Jha, C. M. DesRoches, P. D. Kralovec, and M. S. Joshi, ‘‘A progress report on electronic health records in US hospitals,’’ Health Affairs, vol. 29, no. 10, pp. 1951–1957, 2010.

[4] M. B. Buntin, M. F. Burke, M. C. Hoaglin, and D. Blumenthal, ‘‘The benefits of health information technology: A review of the recent literature shows predominantly positive results,’’ Health Affairs, vol. 30, no. 3, pp. 464–471, 2017.

[5] A. K. Jha, D. Doolan, D. Grandt, T. Scott, and D. W. Bates, ‘‘The use of health information technology in seven nations,’’ Int. J. Med. Inform., vol. 77, no. 12, pp. 848–854, 2008.

[6] M. Liao, B. Shi, and X. Bai, ‘‘TextBoxes++: A single-shot oriented scene text detector,’’ IEEE Trans. Image Process., vol. 27, no. 8, pp. 3676–3690, Aug. 2018.

[7] P. Lyu, M. Liao, C. Yao, W. Wu, and X. Bai, ‘‘Mask textspotter: An endto-end trainable neural network for spotting text with arbitrary shapes,’’ in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 67–83.

[8] A. Mosleh, N. Bouguila, and A. B. Hamza, ‘‘Image text detection using a bandlet-based edge detector and stroke width transform,’’ in Proc. Brit. Mach. Vis. Conf., 2012, pp. 1–12

[9] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[10] Q. Ye and D. Doermann, ‘‘Text detection and recognition in imagery: A survey,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 7, pp. 1480–1500, Jul. 2015.

[11] X. Liu, D. Liang, S. Yan, D. Chen, Y. Qiao, and J. Yan, ‘‘FOTS: Fast oriented text spotting with a unified network,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 5676–5685.

[12] A. Rana and G. S. Lehal, ‘‘Offline urdu OCR using ligature-based segmentation for Nastaliq script,’’ Indian J. Sci. Technol., vol. 8, no. 35, pp. 1–9, 2015. [4] A. Raza, I. Siddiqi, C. Djeddi, and A. Ennaji, ‘‘Multilingual artificial text detection using a cascade of transforms,’’ in Proc. 12th Int. Conf. Document Anal. Recognit., Aug. 2013, pp. 309–313.

[13] S. Unar, A. H. Jalbani, M. M. Jawaid, M. Shaikh, and A. A. Chandio, ‘‘Artificial urdu text detection and localization from individual video frames,’’ Mehran Univ. Res. J. Eng. Technol., vol. 37, no. 2, pp. 429–438, 2018.

[14] A. Mirza, M. Fayyaz, Z. Seher, and I. Siddiqi, ‘‘Urdu caption text detection using textural features,’’ in Proc. 2nd Medit. Conf. Pattern Recognit. Artif. Intell., 2018, pp. 70–75.

[15] ICDAR2017 Competition on Multi-Lingual Scene Text Detection and Script Identification. Accessed: Aug. 2, 2017. [Online]. Available: http://rrc.cvc.uab.es/?ch=8

[16] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, F. Shafait, S. Uchida, and E. Valveny, ‘‘ICDAR 2015 competition on robust reading,’’ in Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR), Tunis, Tunisia, Aug. 2015, pp. 1156–1160.

[17] C. Yao. MSRA Text Detection 500 Database (MSRA-TD500). Accessed: Aug. 2018 [Online]. Available: http://www.iapr-tc11.org/ mediawiki/ index.php/MSRA_Text_Detection_500_Database_(MSRATD500)

[18] X. Ren, Y. Zhou, Z. Huang, J. Sun, X. Yang, and K. Chen, ‘‘A novel text structure feature extractor for chinese scene text detection and recognition,’’ IEEE Access, vol. 5, pp. 3193–3204, 2017.