Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)
IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

AN EFFICIENT OBJECT DETECTION MODEL USING CONVOLUTION NEURAL NETWORKS

Dr. Ulagamuthalvi [1] ,J.B. Janet Felicita[2] ,Abinaya D[3]
[1] Associate Professsor, Dept of CSE, Sathyabama Institute of Science & Technology, India
[2] Student, Dept of CSE, Sathyabama Institute of Science and Technology, Chennai, India
[3] Student, Dept of CSE, Sathyabama Institute of Science and Technology, Chennai, India

Abstract **— Image processing and computer vision have gained an enormous advance in the field of machine learning techniques. Some of the major research areas within machine learning are Object detection and Scene Recognition. Though there are numerous existing works related to the specified fields object detection still encounters numerous challenges when it comes to implementing in the real-time scenario. The problem occurs in the detection due to various objects present in the background. Object detection mechanism detects a specified object when a particular scene is given. Classifiers like SVM and Neural Networks are used to train the classifier in such a way they are able to detect an object when a new image is given. In this paper, we have proposed a model which detects texts from an image. Bounding boxes are used to detect the texts and localize it. The neural network is used to train the model where numerous images having texts are given as the training set. The performance evaluation is done on the model and it is observed that it detects the texts when a new image is given. Object detection is a fundamental problem in computer vision, which aims to detect general objects in images.**

Keywords— ***Object Detection, Bounding Boxes, SVM, Neural Networks, Classification, Texts, Prediction***

I. INTRODUCTION

Numerous research works are taking place across the globe in various fields. One such field is Machine Learning techniques that are catching up a wide variety of research works especially in object detection and scene recognition [1]. Object detection plays a vital role in detecting and identifying objects in a scene or an image such as it is used in identifying the number of vehicles in a particular time on a road [2] or the number of passengers standing in a bus stop during a peak time [3] and a lot more. Whereas scene recognition plays a very immersive role when it comes to classifying flood-affected areas [4] and the land cover classification [5] in a particular region. The detection is done using various classifiers such as SVM classifier [6], KNN classifier [7] and now a day's Neural Networks are playing a huge role in detecting the objects [8].

Object detection is done on an image where the objects are first trained on a various group of objects and then the classifier comes to a stage where when a new image is given it can automatically detect all the trained objects which more accuracy by bounding boxes around it and naming the label with which it had been trained. Previous studies have shown that the use of SVM classifier in detecting objects and bounding it with text boxes [9]. Numerous researchers have even use Convolution Neural Networks to extract the features of the trained objects and use these feature to classify the new objects [10].

In this paper, we have proposed a model that can classify texts from an image. Most of the images contain text within itself and our work detects the texts from a given image and bounds a box around it such that it is easy for the observer. The images are trained with Convolution Neural Networks which deeply extracts all the feature of the image at every layer and performs the computation and gives the output to the next layer. The efficiency of the classifier is observed and it is seen that it can efficiently classify the textual part of an image with less computational time. The rest of the section is as follows: Section II consists of Literature Survey, section III consists of the methodology used in the paper and section III consists of various results obtained. The paper is concluded in the last by mentioning the relevant future works that could be applied or added to the proposed work.

II. Related Work

Many researchers are working on object detection and scene recognition for various applications. In [11], YOLO approach is developed for object detection that bounds the objects with a box and labels the name of the object that it has identified. It is stated that it is more accurate when compared to traditional systems. Text recognition is finding a new pavement in machine learning techniques. Gomez et al. have proposed a selective search algorithm for identifying the texts in a word document. The algorithm generates a hierarchy of word hypothesis and produces an excellent recall rate when compared to other algorithms [12]. Experiments conducted on ICDAR benchmarks demonstrated that the novel method proposed in [13] extracted the textual scenes from the natural scenes in a more efficient way. When compared to conventional approaches, the proposed algorithm showed stronger adaptability to texts in challenging scenarios. Convolutional Neural Networks used in Biomedical Image Segmentation enable precise localization of neuronal structures when observed in an electron microscopic stacks. The selective search algorithm for detecting the objects was used to generate possible object locations. The Selective Search algorithm resulted in a small set of data-driven, class-independent, high-quality locations, yielding 99% recall and a Mean Average Best Overlap of 0.879 at 10,097 locations [14]. Gupta A [15], generated a series of synthetic data that was scalable and very fast. These synthetic images were fully used to train a Fully Convolution Regression Network that efficiently performed text detection and bounding-box regression at all locations and multiple scales in an image. The end model was able to detect the texts in the network significantly out claimed that it outperforms current methods for text detection in natural images. It achieved an F-measure of 84.2% on the standard ICDAR 2013 benchmark. Furthermore, it processed 15 images per second on a GPU.

TextBoxes++ as a single shot Oriented Scene Text Detector which detects arbitrary oriented scene text with both high accuracy and efficiency in a single network forward pass. There was no post-processing process and also had an efficient non-maximum suppression [16]. The proposed model was evaluated on four public data sets. In all experiments, TextBoxes++ claimed that it outperformed competing methods in terms of text localization, accuracy, and runtime. TextBoxes++ achieved an F-measure of 0.817 at 11.6 frames/s for 1024 × 1024 ICDAR 2015 incidental text images and an F-measure of 0.5591 at 19.8 frames/s for 768 × 768 COCO-Text images. A novel approach was proposed in [17] named Cascaded Localization Network (CLN) that joined two customized convolution nets and used it to detect the guide panels and the scene text from a coarse-to-fine manner. The network had a popular character-wise text saliency detection and was replaced with string-wise text

region detection, that avoided numerous bottom-up processing steps such as character clustering and segmentation. Instead of using the unsuited symbol-based traffic sign datasets, a challenging Traffic Guide Panel dataset was collected to train and evaluate the proposed framework. Experimental results demonstrated that the proposed framework outperformed multiple recently published text spotting frameworks in real highway scenarios [18]. Guo et al. proposed a cost-optimized approach for text line detection that worked well for documents that were captured with flat-bed and sheet-fed scanners, mobile phone cameras, and with other general imaging assets [19].

III. Proposed Approach

The system proposed in the current research makes use of Convolution Neural Networks to automatically detect the textual part when present in an image. The dataset is provided to the network and is trained with the Convolution Neural Network. The features of the images are extracted by each and every layer of the network which is used for further processing. In Convolution Neural Network the output of a particular layer acts as the input of the next corresponding layer. Fig 1. shows the overall architecture of how Convolution Neural Networks work. When an image is given the text regions are extracted and is passed on to the recognition network where the features are extracted and the labeling of the objects is done. The text regions are then identified and BBoxes are used to bound the textual region of the image. The output of the current layer acts as the input of the next layer. The number of layers used in the network entirely depends on the working of the model. The number of layers and the activation functions could be changed and observed to know the best efficient model.

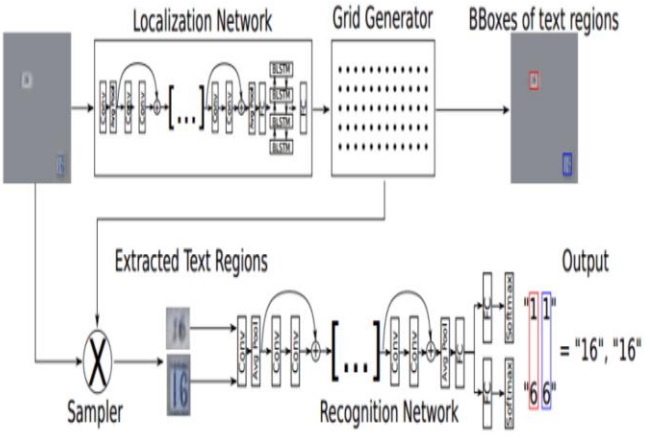

Fig. 1 Architectural Design of Convolution Neural Network in Text Detection

The software design of the proposed model is depicted in Fig. 2. The image shows that the convolutional Neural Networks consists of numerous layers and is fully dependent on the requirement of the application. The image is given as the input and is passed through the neural network.

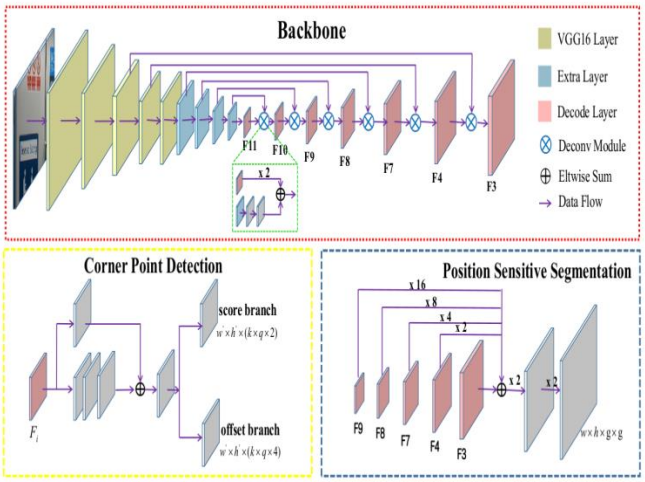

Fig. 2 Software Design

It undergoes a various process like corner edge detection and position sensitive segmentation. This segmentation is used for detecting the textual patterns in the image. Once the network is continuously trained on numerous images, it stores the unique patterns that are

generated when a textual part is present in the image. After the training is done, the testing part comes where a new image is given to the classifier to detect the textual part present in the image. The training set consists of 70% of images whereas the testing set consists of 30% of images. The overall flowchart of the model is described in

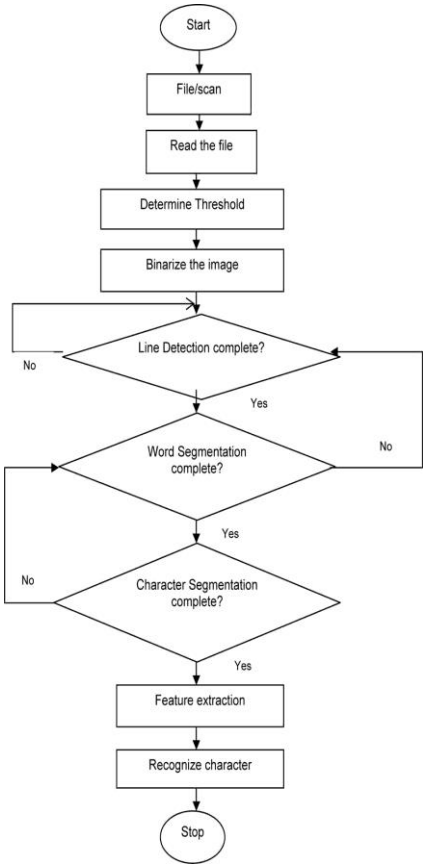

Fig. 3 Flow Chart of the Proposed Model

Fig. 3. The image is first scanned and read from the dataset. Then the image is processed in such a way that the quality of the image is increased s that it could contain more spatial and specification details. Once the image is converted the features are extracted by various layers in the network. Once the training is done, a new image is given to the classifier which reads the image. Once the image is read, the outline of the image is detected first. Then the word segmentation process takes place where the words in the image are segmented separately and then the character recognition process takes place where each character in the text is segmented. This gives the total characters in the image. Then a bounding box is bounded around the textual patterns that are identified by the classifier.

IV. EXPERIMENTAL RESULTS

The experimental results were done on various datasets. The dataset consisted of various images that contained texts within it. The experiment was performed in MAT Lab R2017b where a convolution neural Network was built. The Computer Vision toolbox present in the MAT Lab was used to generate the network. Once the network was created using various layers, the dataset was loaded to perform the training. The training was given on the classifier where the features were extracted. Then a new image was given to the classifier to detect the texts from the images given. Fig. 4 shows the experimental results that were obtained when a new image was given. The image consisted of both a textual part and a non-textual part. The performance of the classifier was very accurate and the results are shown in the following figure. The classifier was able to detect the textual part in the image and bounded a box around the text to make it visible to the observer. well by correctly

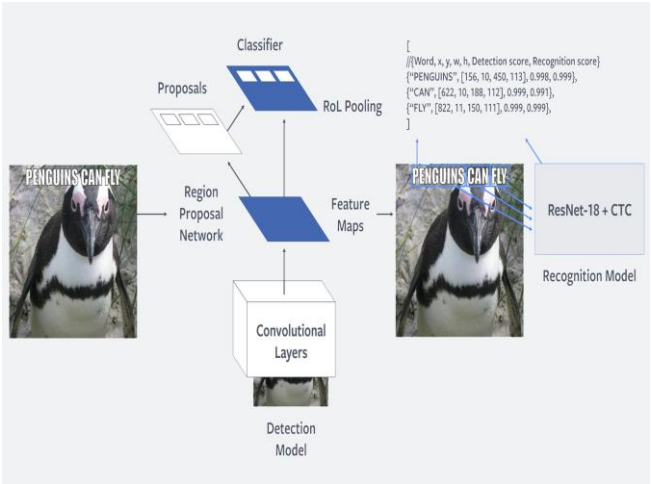

Fig. 4 Implementation Setup

IV. CONCLUSION

Object Detection and scene recognition have gained rapid growth in the upcoming technical era. Most of the applications depend on object detection and scene recognition like the number of areas affected by the flood and the extent of land cover in the country. In our proposed work, we have proposed a model that could detect the texts from the given images and separate it by bounding boxes. The classification was performed using Convolution Neural Networks and the proposed model had good accuracy. Various other parameters such as efficiency and computational time for the process was also considered which was found to be better than the traditional methods. The Future works may include the use of other deep learning techniques and also working on the security concerns of the neural networks.

References

[1] Pandey, M., & Lazebnik, S. (2011). Scene recognition and weakly supervised object localization with deformable part-based models.

[2] Pawlicki, J. A., Mcmahon, M. A., Chinn, S. G., & Gibson, J. S. (2006). *U.S. Patent No. 7,038,577*. Washington, DC: U.S. Patent and Trademark Office.

[3] Castello, T. J., & Tkacik, P. J. (2009). *U.S. Patent No. 7,490,841*. Washington, DC: U.S. Patent and Trademark Office.

[4] Degiorgis, M., Gnecco, G., Gorni, S., Roth, G., Sanguineti, M., & Taramasso, A. C. (2012). Classifiers for the detection of flood-prone areas using remotely sensed elevation data. *Journal of Hydrology*, *470*, 302-315.

[5] Nemmour, H., & Chibani, Y. (2006). Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *61*(2), 125-133.

[6] Osuna, E., Freund, R., & Girosit, F. (1997, June). Training support vector machines: an application to face detection. In *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on* (pp. 130-136). IEEE.

[7] Liu, C. L., Lee, C. H., & Lin, P. M. (2010). A fall detection system using k-nearest neighbor classifier. *Expert systems with applications*, *37*(10), 7174-7181.

[8] Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep neural networks for object detection. In *Advances in neural information processing systems* (pp. 2553-2561).

[9] Shin, C. S., Kim, K. I., Park, M. H., & Kim, H. J. (2000). Support vector machine-based text detection in digital video. In *Neural networks for signal processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop* (Vol. 2, pp. 634-641). IEEE.

[10] Gidaris, S., & Komodakis, N. (2015). Object detection via a multi-region and semantic segmentation-aware CNN model. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1134-1142).

[11] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[12] Gómez, L., & Karatzas, D. (2017). Text proposals: a text-specific selective search algorithm for word spotting in the wild. *Pattern Recognition*, *70*, 60-74.

[13] Zhang, Z., Shen, W., Yao, C., & Bai, X. (2015). Symmetry-based text line detection in natural scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2558-2567).

[14] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and*

computer-assisted intervention (pp. 234-241). Springer, Cham

[15] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision*, *104*(2), 154-171.

[16] Gupta, A., Vedaldi, A., & Zisserman, A. (2016). Synthetic data for text localization in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2315-2324).

[17] Liao, M., Shi, B., & Bai, X. (2018). Textboxes++: A single-shot oriented scene text detector. *IEEE Transactions on Image Processing*, *27*(8), 3676-3690.

[18] Rong, X., Yi, C., & Tian, Y. (2016, October). Recognizing text-based traffic guide panels with cascaded localization network. In *European Conference on Computer Vision* (pp. 109-121). Springer, Cham

Guo, Y., Sun, Y., Bauer, P., Allebach, J. P., & Bouman, C. A. (2015, February). Text line detection based on cost-optimized local text line direction estimation. In *Color Imaging XX: Displaying, Processing, Hardcopy, and Applications* (Vol. 9395, p. 939507). International Society for Optics and Photonics.