Classification of Text and Non-Text from Bilingual Document Images Using Deep Learning Approach

Shivakumar G[1] ,Ravikumar M[2] ,Shivaprasad B J[2]

[1] Assistant professor, Dept. of IT & Computer Application Vignan's University, Guntur, AndhraPradesh-India.
[2] Professor Dept. of Computer Science, Kuvempu University, Jnanasahyadri, Shivamogga, India.
[2] Assistant professor, Dept. of Computer Sci., and Eng., Alvas Institute of Eng., Technol., Mangalore, India.

Abstract: In this work, we have presented an efficient approach for classification of text and non-text document information from real time office documents images printed/handwritten which are bilingual using a deep learning approach i.e., U-net architecture for experimentation purpose. We have created our own dataset containing 2000 document images. Initially pre-processing is applied on the input document images proposed method is compared with other existing methods and obtained accuracy of 99.62% different performance measure i.e., (Specificity, Sensitivity, Precision, F1-Score) used in the experimentation.

Keywords: Document Images; Pre-Processing; Filtering; Segmentation (U-Net).

1. Introduction

An imperative aspect of computer vision is the selection and classification of areas of interest in scanned images of text documents. Many researchers around the world are studying how to convert document images into editable formats. There needs to be a separation of text zones from non-text zones and a correct ordering of them in reading systems. An image can be analyzed to detect/extract/recognize text. For applications including optical character extraction, human-machine input distinction, spam detection, and machine-to-human input differentiation, text recognition and classification in natural images are very significant. Changes in the environment in which images are taken make it difficult for in-text recognition to recognize valuable full text in images. Image text detection identifies locations that contain meaningful whole text in an image. Taking an image in a different area makes it difficult. In analyzing document layouts, it is important to separate text and non-text elements.

The complex structure of the document has limited the quality of separation results despite several approaches. In order for the printed text to be recognized, it must be separated from non-text areas, such as signatures, handwritten text, logos, and other symbols, in order to be accurate. Most research, however, focuses on converting images of documents into the editable text because of the many ways in which this conversion can be used.

Survey of text / non-text separation using various feature classifier combinations.

Documents written by hand are generally unstructured. They generally lack structure, i.e., they lack organization. Due to the lack of a specific layout, handwritten documents appear very chaotic compared to printed documents [4].

Data extraction and retrieval from digital documents have become nearly impossible with the rapid increase in digital documents. There is a need for automated methods. A variety of tools and methods are available to convert digital documents into text that can be processed. To understand and extract knowledge from documents, graphical elements like tables, figures, and equations are crucial. In the research community, therefore, the detection of these objects from documents has attracted considerable attention. Detecting tables and figures within documents is a challenging problem due to the lack of common dimensions and variations in their layouts. The purpose of this article is to recognize different types of digitally generated documents that contain graphical objects such as tables, diagrams, and equations. An object detection problem in a natural scene is conceptually similar to this problem. In rule- based systems, it is difficult to detect irregularly arranged equations, tables, and diagrams [1] – [8].

We present an end-to-end deep learning-based framework, called Visual Structure Object Recognition (VSOR), for detecting visual objects in document images, such as tables, figures, and