Intelligent Document Processing Through Deep Learning-Based Document Image Layout Analysis and Retention

Dr. K. Satyam[1], Appineni Abitha[2]

[1]Associate Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India.

[2] Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India.

Abstract

Text blocks, tables, headers, and images are all part of the intricate structural layouts found in document visuals like bills, forms, and reports. For effective document processing and information retrieval, it is crucial to extract and comprehend these layout elements. Because of differences in format, orientation, and content distribution, traditional document analysis approaches frequently fail to correctly detect and maintain the structural layout of documents. This research proposes a deep learning-based method for document picture layout analysis and retention in order to overcome these difficulties. Through a web-based interface, the suggested system allows users to upload document images. A deep learning model is then used to identify and categorize different layout elements, including text, tables, images, and headers.

After processing the supplied image, the system produces an annotated output with bounding boxes highlighting the layout elements that were found. This method increases the accuracy of layout detection while maintaining the document's hierarchical organization.

In order to automatically analyze document structures and display the findings via an interactive online interface, the created system combines deep learning models with image processing techniques. The suggested approach successfully identifies important layout components and facilitates intelligent document processing, according to experimental results. Applications like digital archiving systems, automated document management, and invoice processing can benefit from the system*.*

Keywords

Document Image Processing, Deep Learning, Document Layout Analysis, Object Detection, Intelligent Document Processing, Image Segmentation, Layout Detection

I. Introduction

Automated document processing systems have been increasingly popular in recent years due to the explosive increase of digital documents, such as invoices, reports, forms, and receipts. Businesses and organizations deal with a lot of document pictures that include important information arranged in various layout formats, including text blocks, tables, headers, and photographs. Manually extracting important information from these documents takes a lot of effort and is prone to human mistakes. As a result, computer vision and machine learning have made automated document analysis a significant area of study.

One of the most important procedures in document understanding systems is Document Image Layout Analysis (DILA). Its main objective is to recognize and separate various structural elements in a document image. In order to identify layout elements, traditional document analysis approaches used handmade features and rule-based procedures. Nevertheless,

these methods sometimes fall short when handling intricate document structures, a variety of font styles, diverse formats, and noisy backdrops.

Document layout analysis is now more precise and efficient due to the development of deep learning techniques, especially convolutional neural networks and object detection models.

Compared to conventional methods, deep learning models can more accurately identify layout elements, including text sections, tables, graphics, and headers, by automatically learning features from document images. A deep learning-based system for document picture layout analysis and retention is presented in this paper. Through a web-based interface, the suggested system enables users to upload document images, which are then processed by a trained deep learning model to identify various layout elements. Bounding boxes are used by the system to recognize and highlight certain elements, making the document structure easy to see. While facilitating automatic document understanding, the processed output aids in maintaining the layout information.

II. System Architecture

The suggested document image layout analysis system's overall process is described in the system architecture. The proposed method uses deep learning techniques to automatically identify and preserve layout elements from document pictures. The architecture is made up of multiple modules that cooperate to process the input image and produce the final output with layout elements that have been identified. The user submits a document image via the web interface via the image upload module, which starts the procedure. A variety of layout structures, including text blocks, tables, headers, and images, may be present in the uploaded image. The image is transferred to the backend server for additional processing after it has been posted.

The system gets the input image ready for analysis during the image preprocessing phase. In this step, the image may be resized, pixel values may be normalized, and the image may be formatted appropriately for the deep learning model. Preprocessing guarantees higher detection performance and enhances the quality of the supplied data. The deep learning model in charge of identifying document layout elements receives the image after preprocessing. The trained model examines the document image and recognizes several structural elements, including headers, tables, images, and text areas. To find these components and ascertain their locations within the image, the model employs object detection techniques.

The system creates bounding boxes around each detected element after the layout components have been identified. These bounding boxes categorize the identified regions based on their category and visually emphasize them. This facilitates comprehension of the document's organization and helps maintain its structural layout. Lastly, the processed image and the original image with highlighted layout elements are shown by the result visualization module. Users may clearly examine the detected layout elements thanks to the online interface that presents the output. Additionally, the system offers the ability to upload an additional photograph for additional analysis.

III. Proposed Methodology

The suggested approach focuses on applying deep learning techniques to automatically identify and maintain document layout hierarchies. Image input, preprocessing, deep learning-based detection, and output visualization are all steps in the system's processing of document images. Together, these steps enable the identification and highlighting of various layout components included in a document.

Step 1: Document Image Upload

Uploading the document image via a web-based interface is the system's initial step. Images of forms, reports, invoices, and other document types can be chosen and uploaded by users. After being saved on the server, the uploaded image is sent to the processing module for additional examination.

Step 2: Image Preprocessing

The submitted image is preprocessed before the deep learning model is used. By carrying out tasks including image scaling, normalization, and format conversion, this stage gets the image ready for precise analysis. Preprocessing guarantees compliance with the deep learning model and enhances the quality of the input data.

Step 3: Deep Learning-Based Layout Detection

The trained deep learning model receives the document picture after preprocessing. The model examines the document's visual elements and identifies several layout elements in the picture. Text areas, tables, graphics, and headers are a few examples of these elements. Each layout element's placement is determined by the model using object detection techniques.

Step 4: Bounding Box Generation

The system creates bounding boxes around each identified layout element when it has been identified. The layout elements' locations within the document image are shown by these bounding boxes. The recognized element type, such as text, table, or image, is labeled on each bounding box.

Step 5: Layout Retention and Output Display

In the last step, the system highlights identified elements in the document image to maintain the layout structure. The original image and the processed image with bounding boxes are shown on the online interface. This makes it simple for users to see and comprehend the document structure.

IV. Results and Discussion

The proposed system was tested using different document images to evaluate its ability to detect and preserve document layout structures. The system provides a user-friendly interface that allows users to upload document images and view the detected layout components. The experimental results demonstrate that the system can successfully identify and highlight various layout elements present in the document.

Figure 1. Document Image upload Interface

The suggested system's document upload interface is depicted in Figure 1. Users can upload a document image for layout examination on this page. Users can select the necessary document image from their local system using the interface's file selection feature. The user can submit the image to the system for processing after choosing it.

In order to facilitate users' interaction with the system, the interface is made to be straightforward and easy to use. After the image is uploaded, it is transferred to the backend server, where the deep learning model analyzes it and identifies the layout elements.

Figure 2: Document Layout Detection Output

The result produced by the suggested document layout detection method is shown in Figure 2. After processing the supplied document image, the system recognizes several layout elements, including text areas, tables, photos, and headers.

Colored bounding boxes, which provide the precise location of each layout element within the document, are used to emphasize the discovered elements. The type of detected element is indicated by the labels attached to the bounding boxes. For instance, the system might identify sections of the document, such as the picture, table, text, and header.

This graphic aid makes it easier to comprehend the document's hierarchical organization. The outcomes show that the deep learning model can reliably identify various layout components while maintaining the document image's layout structure.

V. Conclusion

A deep learning-based method for document picture layout analysis and retention is shown by the suggested system. From document pictures, it automatically identifies structural elements, including text, tables, graphics, and headers. Through the system's web-based interface, users can upload documents and use bounding boxes to see layout elements that have been identified. The model successfully recognizes document structures and maintains layout information, according to experimental results. Document management, data extraction, and digital archiving are examples of automated document processing applications that can be supported by this method.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[3] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.

[4] A. K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames," Pattern Recognition, vol. 31, no. 12, pp. 2055–2076, 1998.

[5] D. Doermann, "The Indexing and Retrieval of Document Images: A Survey," Computer Vision and Image Understanding, vol. 70, no. 3, pp. 287–298, 1998.