

Abstract

With the emergence of natural language processing and deep learning, the idea of extracting key-fields from invoices is gaining more prominence in the business sector. However, there is lack of work and common understanding regarding this topic in scientific literature. It is known that financial reimbursement is considered as a burdensome process in extracting named entities from real time invoices. Besides, considering the resolution of images into account whether its clear enough to read the text or not. Because, most of the invoices are in paper format, they either need to be scanned or a photo of it is taken by the camera and the result of it is not always a good quality. So, to solve these problems a method like Chargrid came into existence, which is a recently proposed approach of understanding 2-dimensional documents. Here, the document is represented in grid format along with its spatial structure to process it to the model. Each character grid from the document is embedded into one-hot encoding. Further, this character grid in image format is passed as an input to Convolutional neural network (CNN). This thesis work aims to automate the invoice extraction process with the integration of deep learning algorithms. An attempt was also made to implement invoice key-field extraction using object detection model YOLOv3 to accurately locate, segment and intercept the key-field areas on invoices apart from chargrid-FasterRCNN. The base networks used for these models are Darknet-53 and VGG-16 and results of the Chargrid-FRCNN are then compared with YOLOv3 model. A brief analysis is given on the performance of chargrid-FRCNN along with possible solutions to improve the detection accuracy further. Finally, a RPA tool is integrated with above mentioned algorithms to automate the workflow of invoice process.

Keywords: 2D document understanding, embedding, Object detection, Tesseract-OCR, Robotic process automation