Abstract

Font family and font size recognition became an essential step for document analysis. Font recognition helps to identify the proper segmentation method to be used before feeding the document to the Optical Character Recognition (OCR). In this paper, some of the previous techniques used for font family and font size recognition will be discussed then we will discuss the proposed method that is based on deep learning. Two methods have been presented in this paper 1) a method for font family recognition (font size invariant) and 2) a method for font size recognition. Both methods use Deep Convolutional Neural Networks (D-CNN). We evaluated the proposed method on Arabic Printed Text Image Database (APTI) [7] and on a document generated using APTI database word images and scanned with the scanner.

References

Deep Arabic Font Family and Font Size Recognition

7. APTI Arabic Printed Text Image Database. https: //diuf.unifr.ch/diva/APTI/index.html

Index Terms

Computer Science

Information Sciences

Keywords

Font Family Recognition, Font Size Recognition, Optical Character Recognition (OCR), Document Layout Analysis (DLA), Deep Learning, Deep Convolutional Neural Network (D-CNN)