

Small Text Extraction from Documents and Chart Images

Authors : Rominkumar Busa, Shahira K. C., Lijiya A.

Abstract : Text recognition is an important area in computer vision which deals with detecting and recognising text from an image. The Optical Character Recognition (OCR) is a saturated area these days and with very good text recognition accuracy. However the same OCR methods when applied on text with small font sizes like the text data of chart images, the recognition rate is less than 30%. In this work, aims to extract small text in images using the deep learning model, CRNN with CTC loss. The text recognition accuracy is found to improve by applying image enhancement by super resolution prior to CRNN model. We also observe the text recognition rate further increases by 18% by applying the proposed method, which involves super resolution and character segmentation followed by CRNN with CTC loss. The efficiency of the proposed method shows that further pre-processing on chart image text and other small text images will improve the accuracy further, thereby helping text extraction from chart images.

Keywords : small text extraction, OCR, scene text recognition, CRNN

Conference Title : ICDAR 2022 : International Conference on Document Analysis and Recognition

Conference Location : Jerusalem, Israel

Conference Dates : November 29-30, 2022