Text Localization in Fixed-Layout Documents Using Convolutional Networks in a Coarse-to-Fine Manner

Authors : Beier Zhu, Rui Zhang, Qi Song

Abstract : Text contained within fixed-layout documents can be of great semantic value and so requires a high localization accuracy, such as ID cards, invoices, cheques, and passports. Recently, algorithms based on deep convolutional networks achieve high performance on text detection tasks. However, for text localization in fixed-layout documents, such algorithms detect word bounding boxes individually, which ignores the layout information. This paper presents a novel architecture built on convolutional neural networks (CNNs). A global text localization network and a regional bounding-box regression network are introduced to tackle the problem in a coarse-to-fine manner. The text localization network simultaneously locates word bounding points, which takes the layout information into account. The bounding-box regression network inputs the features pooled from arbitrarily sized RoIs and refine the localizations. These two networks share their convolutional features and are trained jointly. A typical type of fixed-layout documents: ID cards, is selected to evaluate the effectiveness of the proposed system. These networks are trained on data cropped from nature scene images, and synthetic data produced by a synthetic text generation engine. Experiments show that our approach locates high accuracy word bounding boxes and achieves state-of-the-art performance.