

Literature survey on text detection in natural images

[Technology](#), [Artificial Intelligence](#)



Zhanzhan Cheng et al. (2018) describes the most challenging problem of text detection in natural images as scene texts are often in irregular (e. g. curved, arbitrarily oriented or seriously distorted) arrangements. This paper mainly focuses on the network which can be trained end-to-end by using only images and word-level annotations. This system uses the arbitrary orientation network, which are combined into an attention-based decoder to generate character sequence. It can be useful for many machine vision applications such as road sign recognition and navigation reading for advanced driver assistant system. This module includes the arbitrary orientation network (AON) and the filter gate (FG), which constitute the core of the proposed method. This method can effectively recognize both irregular and regular texts from images. Experiments over both regular and irregular benchmarks (SVT-Perspective, CUTE80, ICDAR 2015, IIIT5K-Words, Street View Text, ICDAR 2003) validate the superiority of the proposed method. The curves formed by connected red arrows indicate text placement trends. All texts in the images are correctly recognized by our method.

Minghui Liao et al. describes a challenging problem of text detection in natural images. Arbitrary orientations, small sizes, and significantly variant aspect ratios are the main challenges of scene text detection in natural images. This paper mainly focuses on scene text detector, which is fast and end-to-end trainable named Textboxe++, detects arbitrary oriented scene text with both high accuracy and efficiency. The architecture of TextBoxes+++, a fully convolutional network including 13 layers from VGG-16 followed by 10 extra convolutional layers, and 6 Text-box layers connected to 6 intermediate convolutional layers. Paper proposes some special designs to

<https://assignbuster.com/literature-survey-on-text-detection-in-natural-images/>

adapt SSD network for efficiently detecting oriented text in natural images. Specifically, it proposes to represent arbitrary-oriented text by quadrilaterals or oriented rectangles. The whole training process (including the pre-training time on Synthetic Text dataset) takes about 2 days on ICDAR 2015 Incidental Text dataset, which is now-a-days the most tested dataset. They have tested the pipeline of TextBoxes++ followed by CRNN model on three popular word spotting or end-to-end recognition benchmark datasets: ICDAR 2015 Incidental Text dataset, SVT dataset, and ICDAR 2013 dataset. TextBoxes++ achieves an f-measure of 0.817 at 11.6 frames/s for 1024×1024 ICDAR 2015 incidental text images and an f-measure of 0.5591 at 19.8 frames/s for 768×768 COCO-Text images.

Youbao Tang et al. (2018) describes a greatly challenging problem in computer vision which is scene text detection. This paper proposes a new scene text detection method that includes superpixel-based stroke feature transform (SSFT) and deep learning based region classification (DLRC). The proposed method consists of two stages: character region extraction and text region detection and scene text detection approaches, first use CNNs or some detection mechanisms (e. g. , edge boxes) to extract a number of region proposals, and then filter out non-text proposals. The Caffe library is used to implement the deep learning based framework, and all other parts of the proposed method are implemented using MATLAB. The proposed method is evaluated on three publicly available datasets: ICDAR2011, ICDAR2013, and street view text. It achieves F-measures of 0.876, 0.885, and 0.631, respectively, which demonstrate the effectiveness of the proposed scene

text detection method. This method still fails in some cases. For example, when the text and background have very similar intensity or low contrast, or the scene images have low resolution or low quality, the proposed method cannot segment text from background after using the super pixel based clustering strategy.

Xuebo Liu et al. (2018) describes Incidental scene text spotting which is considered one of the most difficult and valuable challenges in the document analysis community. This paper mainly focuses on end-to-end scene text detection. They proposed a unified end-to-end trainable Fast Oriented Text Spotting (FOTS) network for simultaneous detection and recognition, sharing computation and visual information among the two complementary tasks. Architecture of proposed work uses Specially, RoIRotate which is introduced to share convolutional features between detection and recognition. Experiments on ICDAR 2015, ICDAR 2017 MLT, and ICDAR 2013 datasets demonstrate that the proposed method outperforms state-of-the-art methods significantly, which further allows us to develop the first real-time oriented text spotting system which surpasses all previous state-of-the-art results by more than 5% on ICDAR 2015 text spotting task while keeping 22.6 fps.

Ghulam Jillani Ansari et al. (2018) describes extraction of text in images which is popular research field in the computer vision in recent time. This paper focuses on an exigent aspect such as natural scene text detection and extraction, which is investigated due to cluttered background, unstructured scene, orientations, ambiguities and much more. This system consists of LUV

channel, MSER techniques, LBP and T-HOG feature descriptor, SVM classifier and CNN Network. The and uniqueness of proposed work is that it not only recognizes the text intelligently but also removes error in order to preserve actual meaning of the text. The application is programmed in C and Python on Ubuntu platform. The presented CNN is trained with positive and negative samples collected from Char74K, IIIT5K, ICDAR2003 and SVT datasets. This work has been evaluated on 3 standard dataset and finds out that is perform well on all standard dataset.

Yuanwang Wei et al. (2018) describes about the detection of text from scene image which is the essential part of many real-world application. This paper mainly focuses on an algorithm which is designed to prune the edges that do not represent an adjacency relation of the candidate regions. In this paper an effective coarse-to-fine algorithm is proposed to prune the non-adjacent edges (those in which two corresponding characters are not adjacent to each other in the same word). This paper contributes in (1) designing a parallel processing architecture which integrates non character region filtering with candidate generation, and (2) At the text-candidate construction stage, argument regarding the text candidate construction problem can be posed as one of splitting an undirected, fully connected graph into connected sub-graphs. So presented a novel multi-oriented scene text detection method that employs a CNN model to filter out the non-character regions and groups characterc and candidates into text lines based on pruning non-adjacent edges from a graph. Three differently oriented public datasets were employed to evaluate the proposed method: the Oriented Scene Text

Dataset (OSTD) and the USTB Street View Text Detection 1000 Database (USTB-SV1K). They conducted all the experiments on a computer with an Intel(R) Core(TM) i72. 4GHz4-core CPU with 16GB of RAM running a Windows64-bit operating system). At run time, all the testing images whose widths were greater than 1024 pixels were rescaled to a width of 1,024 pixels, while the aspect ratio was kept unchanged. After size normalization, (including text localization) is approximately 2.1s for each image.

Shaohui Ruan et al. (2018) describes the most challenging problem of scene text detection called arbitrarily oriented text. This paper mainly focuses on prediction of word level bounding boxes via fully connected network. They proposed a method which extracts the feature from the input image by residual network and apply multi-level fusion over the extracted features. The pipeline consists of a fully convolutional network and a standard NMS as post-processing. This method achieves an F-measure of 83.46% and 56.39% on ICDAR 2015 Incidental Scene Text benchmark and COCO-Text dataset respectively, outperforming the previous methods by a large margin. Also, it can run at over 11 FPS on 704×1280 images, which is much faster than the previous works.

Wafa Khelif et al. (2018) describes reading of text embedded in natural scene text detection which is essential for many application. In this paper, They proposed a method for detecting text in scene images based on multi-level connected component (CC) analysis and learning text component features via convolutional neural networks (CNN), followed by a graph-based grouping

of overlapping text boxes. The system is evaluated on the standard public dataset of the ICDAR2013.