Natural Scene Text Detection Algorithm Based on Improved DBNet

Chen Hui Yang, Liu Jing, Zhou Wei Min
Department of Automation, Shanghai University

Abstract

There are many problems in text detection, such as large scale differences of high-resolution image features and poor multi-scale feature fusion, we propose an improved algorithm based on dbnet. On the basis of the feature fusion module, we add a atrous Convolution network with kernel-shared pooling to increase the receptive field, so that higher-level semantic information can be obtained in the feature fusion network, and through the shared kernel, the number of model parameters can be reduced, the computational cost can be reduced, and the detection accuracy can be improved. At the same time, we add the attention mechanism into the residual network to suppress the complex background noise and promote the information interaction between channels; In the loss function, we use dice loss partially used to solve the imbalance of positive and negative sample data. Our experimental evaluation is on ICDAR2013 and ICDAR2015 datasets. The experimental results show that the algorithm has a certain improvement in accuracy and F value.

Algorithm Analysis

The overall structure of this method is shown in Figure 1. The feature extraction network takes ResNet-50 as the backbone network. At the same time, the channel attention mechanism SENet is used in the backbone network to enhance the feature interaction between channels, so as to enhance the ability of network feature extraction. Through the kernel-shared pooling atrous Convolution and feature pyramid network KPA-FPN, the receptive field is expanded, the information of different scales is obtained, and the recognition ability of features is improved. Then the feature pyramid performs up sampling from top to bottom, cascading the up sampled features and features with the same size to obtain the fused feature map.The feature map passes through two predictors respectively, and 2 predictors generate prediction probability map and prediction threshold map respectively. Finally, the generated probability map and threshold map through a binarization module to generate the final binarization image.

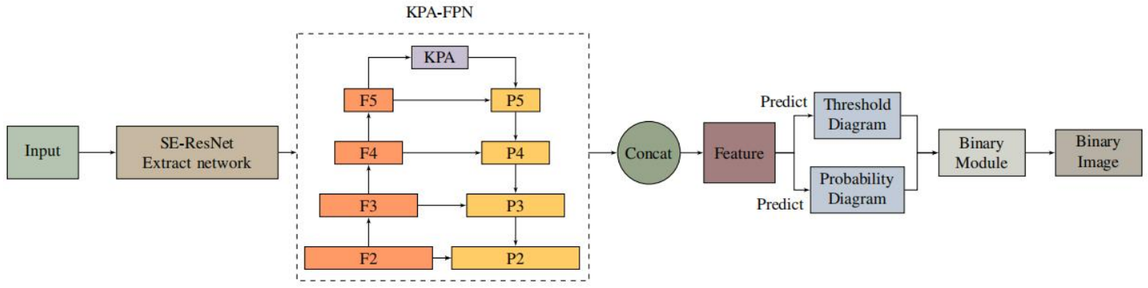

Figure 1. Overall structure

Among them, the kPa module is a further improved structure based on ASPP (atlas spatial), which adds the shared convolution core to the cascade atrus convolution network and its kernel sharing and merging atrophic convolution module. As shown in Figure 2.

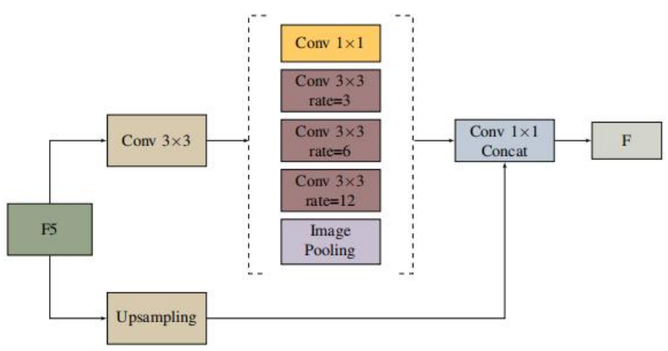

Figure 2. Network structure of KPA

Experimental Results

ICDAR2015 and ICDAR2013 are used in the data set of this experiment. The evaluation indexes used in this experiment are Precision, Recall and F-measure. Generally, precision refers to the proportion of positive samples in the number of detected samples; Recall rate refers to the proportion of the number of positive samples detected in all positive samples, due to the restrictive relationship between accuracy and recall, F- measure is introduced as the balance index between them.

Table 1 shows the experimental comparison of this algorithm and DB algorithm in dataset ICDAR2013.

Table 1 ICDAR2013 dataset algorithm comparison

Method	P(%)	R(%)	F(%)
DB	86.41	78.72	82.39
Ours	88.17	79.53	83.62

According to the analysis of the experimental results in Table 1, the improved algorithm that we use has better detection results when processing horizontal text, in which the accuracy P is increased by 1.76% compared with the basic algorithm, the recall rate R is increased by 0.81%, and the F value is increased by 1.23%, indicating that the algorithm in that we use is better than the improved algorithm for horizontal text detection in complex natural scenes.

Table 2 shows the experimental comparison of this algorithm and DB algorithm in data set ICDAR2015.

Table 2 ICDAR2015 dataset algorithm comparison

Method	P(%)	R(%)	F(%)
DB	91.31	80.32	85.46
Ours	92.16	81.43	86.46

From table 2, we can see that the accuracy P of the algorithm that we use has increased by 0.85%, the R rate has increased by 1.11%, and the F value has increased by 1%, indicating that the algorithm that we use has a good effect on oblique text detection in complex natural scenes, the algorithm that we use has been greatly improved.

Conclusion

In order to solve the problems of text detection in natural scenes, we propose a text detection method based on kernel- shared pooling atrous convolution and improved loss function. In the feature extraction stage, we add SE attention mechanism into the backbone network to suppress complex background noise and get more accurate feature map. In the feature fusion stage, we propose an improved model of KPA -FPN to increase the receptive field and obtain multi-scale semantic information. Secondly, we use dice loss function partially in the loss function to correct the imbalance of positive and negative samples. The experimental results show that the detection method that we use is feasible and effective, but there are still problems of high computational complexity and insufficient differentiation between text region and non text region, which will be further improved in the follow-up research work.