Convolution Neural Ruler: A Measurement Method and Application in Multi-Oriented Text Detection

**Abstract**

In this paper, we propose a method called ‘Convolution Neural Ruler’ (CNR) to measure distance by summing values of different dimensions which is just like using a ruler. To demonstrate the efficiency of CNR, we apply it into a non-proposal based detection system which is used to detect multi-oriented scene texts with various sizes, orientations and aspect ratios. On benchmark ICDAR2015 which focuses on incidental texts, the F1-measure achieves 77% surpassing the second place by a large extent. We also give convincing results on benchmark ICDAR2013 and MSRA-TD500.

**Introduction**

Recently, convolution neural networks (CNNs) have significantly driven the improvements of object detection. Unlike object classification or segmentation which only perform classification on image-level or pixel-level, object detection requires not only to recognize the object, but also to measure the bound of each object. As a result, most CNN based detection structures like [Fast RCNN] [Faster RCNN] [SSD] [YOLO] [Densebox] are multi-task structure with one recognizer for classification and one regressor for localization.

According to the design of regression task, we can divide recent detection methods into two groups: proposal based regression method [Fast RCNN] [Faster RCNN] [SSD] and non-proposal based regression [YOLO] [Densebox] method. The former predicts the offset from the proposal to the ground truth and the latter directly predicts the bound of an object. Current high performance detection structures like [Faster RCNN] [SSD] are belong to proposal based regression method and they benefit from the easier regression task in which proposals are not far from the corresponded ground truth. On the contrary, the non-proposal based regression method struggles to localize objects correctly which has been pointed out in [YOLO].

Despite proposal based regression is superior to non-proposal based one, it cannot avoid the problem that it will be difficult to get ideal results if we have to output a bounding box whose aspect ratio is too large and this will be normally encountered in detecting text lines [See Fig.1]. Simply increasing the diversity of ‘anchors’ like [Deep-Text] may be effective but sacrifices efficiency of the whole detection system.

The reason why non-proposal based regression struggles to localize object correctly, as well as the non-proposal based regression fail to localize object with too large aspect ratio, is that it is difficult to directly regress a distribution with large variance by using the Euclidean Loss. And it is also the case even for the toy problem shown in Fig.2.

To solve this problem, we propose the ‘Convolution Neural Ruler’ (CNR) to measure the distance like a ruler. First we set several dimensions like kilometer, meter and millimeter. Then we regress values ranging from zero to one for each dimension. Finally we combine the values of each dimension to get the predicted result. The whole process is just like how we use a rule --- we actually measure the length by adding values of each dimension.

To test the efficiency of ‘CNR’, we design a non-proposal based detection system and choose scene texts whose aspect ratios and orientations vary much more than general objects as the detection targets. The results, especially for long and multi-oriented texts, given by our method demonstrate the reasonableness and efficiency of the proposed method.

To sum up, the main contributions of this paper are in three folds: firstly, we propose a method called ‘Convolution Neural Ruler’ to make it possible directly regress arbitrary values; secondly, we propose a non-proposal based detection system to localize multi-oriented scene text with CNR embedded to accurately determine the bounds of text. Benefitting from the precise bound given by our system, we have got state of the art result on benchmarks with multi-oriented text surpassing recent results with a large margin.

The reminding parts of this paper are arranged as follows: Section 2 gives a brief review of object and scene text detection, Section 3 introduces the details of our detection system including , Section 4 displays the results and analysis of reasonableness of CNR and Section 5 concludes our work.

**Related Work**

**Bounding box regression**

**Scene text detection**

**Convolution Neural Ruler**

System Overview

Convolution Neural Ruler

Asymmetric Euclidean Loss

Re-Adjusted NMS

**Experiments**

Training details

Result

ICDAR2015, MSRA-TD500, ICDAR2013

Analysis

The effectiveness of asymmetric L2 loss

The learned coding strategy

**Conclusion**