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

Anonymous CVPR submission

Paper ID ****

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.

1. 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.

2. Related Work

Deep model based regression

2.0.1 Scene text detection

3. Proposed Method

In this section, we describe the proposed method in detail. First we give an overview of the non-proposal based detection system, Second we formulated 'CNR' embedded in our detection system. Third we

3.1. System overview

3.2. Convolution neural ruler

3.3. Network structure

3.4. Asymmetric regression loss

4. Experiments

5. Conclusion

6. title

6.1. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: http://www.pamitc.org/documents/mermin.pdf.

6.2. Blind review

so prefer [2, 1, 4] to [1, 2, 4].

6.3. Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

References

[1] A. Alpher. Frobnication. *Journal of Foo*, 12(1):234–778, 2002. 2

[2] A. Alpher and J. P. N. Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003. 2

[3] A. Alpher, J. P. N. Fotheringham-Smythe, and G. Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004.

[4] Authors. The frobnicatable foo filter, 2014. Face and Gesture submission ID 324. Supplied as additional material fg324.pdf. 2

[5] Authors. Frobnication tutorial, 2014. Supplied as additional material tr.pdf.

[1] This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

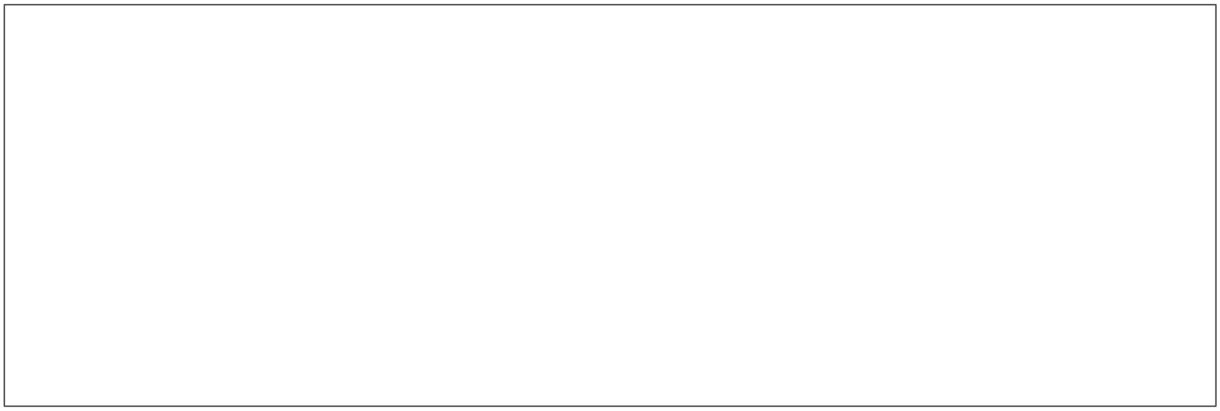

Figure 2. Example of a short caption, which should be centered.