Review for End to End Object Detection with Transformers

Can Koz

April 18, 2021

1 Summary

Existing object detection pipelines use several post processing stages such as anchor boxes or window centers and authors propose an end to end approach to object detection problem by using self attention mechanisms of transformers which outperforms the other methods in sequence prediction. Authors evaluate their technique on COCO dataset against a very competitive FASTER R-CNN baseline. The proposed method can be easily adapted to more complex tasks such as panoptic segmentation.

2 Strengths

- Attention mechanisms and transformers has changed the deep learning algorithms in sequence prediction and authors adapted this approach to object detection problem.
- End-to-end philosophy has led to significant advances in complex structured prediction tasks and authors aim to achieve competitive performance with this approach.
- The flexible design of the method easily extend to complex tasks.

3 Weaknesses

- The technique can be considered as new approach on object detection and as authors mentioned it must be evaluated on diverse task such as small object detection.

4 Evaluation

The method is compared with Faster R-CNN and Retinanet on COCO 2017 dataset for object detection task. In the ablation studies, they tried to explore the contributions of number of encoder layers, number of decoder layers, importance of FFN and importance of positional encodings. In object detection, in terms of AP metrics they have comparable performance to heavily tuned Faster R-CNN baseline.

Additionally, authors evaluate the performance for panoptic segmentation task.

5 Final Comments and Future Work

In the end, transferring transformers, which has revolutionized the NLP domain, to object detection task has result in promising performance. Even though it was published recently, it has attracted the attention of many researchers. Using transformers in the backbone may lead to improvement on small object detections as mentioned in BotNet [Bottleneck Transformers for Visual Recognition] paper.