Object Detection with Transformers Project Proposal

Jazon Samillano

19 December 2020

Overview

The use of Transformers to perform object detection on images is an exciting new endeavor in AI.

1. What is wrong with the current methods of using transformers on object detection?
   a. The average precision (AP) is low relative to other state of the art solutions. For example, on the COCO objection detection test, the use of transformers ranks 17th on the list. This is using Facebook's DETR model.
2. What is your idea in addressing this gap?
   a. Instead of trying to push higher on a generic dataset, I would like to fine-tune Facebook's DETR to become more specialized in a particular domain, such as for self-driving cars.
3. Why do you think your idea will work or make sense?
   a. This idea came about when I tried fine-tuning Google's Faster R-CNN called InceptionResNetV2. I fine-tuned it to start detecting skin cancer, and the results look great. I think I can fine-tune Facebook's DETR to become really good at detecting objects on the road that are critical for the safe operation of self-driving cars.
4. What are the competing ideas on the same problem that you are attacking?
   a. Yolo and Faster R-CNN are really great at real-time object detections. I'm not sure how far transformers will go in this domain, but I'd like to explore it further

just to learn, just in case transformers later on become the gold standard for real-time object detections.

5. Why do you think your idea is better?
   a. The idea of fine-tuning DETR for a specific domain will bypass the very difficult task of becoming the world's number one in generic object detections such as in the COCO dataset.
6. How are we going to measure the performance of your proposed algorithm?
   a. I plan to fine-tune both DETR and Google's Faster R-CNN InceptionResNetV2 on road objects critical to self-driving cars. I will then assess which path proves more promising.
7. Are there available public datasets that you can use? Are there current benchmarks that you can use to compare the performance of your algorithm? If none, are you going to make one?
   a. There are several datasets to choose from, such as the Waymo Open Dataset and the Lyft Level 5 Dataset.
8. What are the metrics that you are going to use to measure the success/failure of your idea?
   a. If I can fine-tune DETR to be a more practical object detector for self-driving cars than Google's InceptionResNetV2, then I would label this a success. However, how will I define "more practical". I am still baffled by that conundrum, and will update this proposal as I get more mental clarity on that issue.