Week 7 paper summary

Sina Mehdinia

05/14/2021

Summary

Redmon et al. [1] invented the YOLO algorithm, which performs real time object detection. YOLO, is very fast comparing to the state of the art of that time, R-CNNs and DPMs and has lots of practical value. This is mainly because there is no pipeline and a single neural network is doing all the tasks. This is inspired by the fact that we humans only look once at an image and detect the objects. The logic behind it is to see the object detection as a regression problem. Each image, is divided into a 7×7 grid cell. And there are only a few bounding boxes predicted for each grid cell (for example 2). There are 5 parameters for each bounding box which are height, width, center coordinates, and confidence, and at last a class probability for each category. There are two versions of the algorithm, YOLO and Fast YOLO. Fast YOLO is the fastest real time object detection of that time with 155 fps. It has 9 layers. YOLO has 24 layers and the architecture is inspired by Inception [2]. The confidence score is the intersection over union (IOU) between the predicted box and and the ground truth. The loss function has multiple parts. It considers if the an object is present in that grid cell and also the bounding box coordinate error if that predictor is responsible for the ground truth box. YOLO uses leaky RELU activation for almost all layers. Comparing to R-CNNs, YOLO is much faster but makes more localization errors, although the errors are much lower in terms of background false positivies. At last, YOLO has a very good performance in learning generalizable representation of the objects.

References

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.