

Project proposal for applied deep learning course.

Course: Applied deep learning

Name: Sina Ahmadi

Student ID: 12104942

Professor: Alexander Pacha

Vienna university of technology

Winter semester 2022

Topic of the project: Object detection

Type of project : This project is a combination of “Bring your own method” and “Beat the classics”.

What I am planning to do is to improve the current deep learning approaches in object detection. I try to implement and improve the current transformer based approaches.

Summary

a) Short description of the project idea and the approach:

As mentioned before, object detection is a relatively new field. I got the idea of the project by comparing the new approaches in object detection with the previous ones. In general, the task of object detection algorithms is to predict a set of bounding boxes and category labels for each object of interest. Former object detection approaches are mostly region based ones. They address this set prediction task in an indirect way, by defining surrogate regression and classification problems on a large set of proposals, anchors, or window centers. Two of the most well-known methods in this regard were Faster RCNN(a two-step approach), and YOLO network(a one step approach). The problem with these approaches were that they needed additional information that should have been provided by the network designers. To simplify these pipelines, we propose a direct set prediction approach to bypass the surrogate tasks. This end-to-end philosophy has led to significant advances in complex structured prediction tasks such as object detection, machine translation or speech recognition. To achieve this, the newer approaches used transformers instead of the complicated neural networks in previous versions. This had some serious advantages. First, it was much easier than designing the complicated neural networks that were needed in previous versions, and furthermore, it wasn't needed to provide knowledge to the algorithm. Actually, the AI itself learned the things it needed. Now I want to dig deeper in the transformer based methods. End-to-End Object Detection with Transformers is the main focus of my project.

b) Description of the dataset collection:

In this project I am going to use the COCO dataset. COCO (Microsoft Common Objects in Context) is a large-scale image dataset containing 328,000 images of everyday objects and humans. The dataset contains annotations you can use to train machine learning models to recognize, label, and describe objects.

c) Different part of the project with time estimation of each part:

Designing and building an appropriate network; training and fine-tuning that network; building an application to present the results; writing the final report; preparing the presentation of your work.

Based on the approach that I have in mind, I need to devote most of time in understanding, implementing, and improving the related neural networks! I am using a popular, famous dataset. So I don't need to spend time on that. My plan for the project can be summarized in the following table:

Task	Time estimation(hours)
Dataset collection	—
Designing and building an appropriate network	80
Training and fine-tuning that network	40
Building an application to present the results	30
Writing the final report	10
Preparing the presentation of your work	3
TOTAL	163

References:

1. S.A.Zaidi, M.Ansari, A.Aslam,N. Kanwal, M. Asghar B.Lee, A Survey of Modern Deep Learning based Object Detection Models, arxiv.2104.11892 (2021).
2. N.Carion, F.Massa, G. Synnaeve, N.Usunier, A. Kirillov,S. Zagoruyko, End-to-End Object Detection with Transformers, arxiv.2005.12872 (2020)