CS323: Deep Learning for Visual Computing

Reading Assignment 4

Due Date - 24/04/2023

Welcome to the fourth reading assignment of CS323. These assignments aim at preparing you for the projects through providing you with important readings. You will be expected to answer questions related to the readings, material provided in class, or both. Good Luck ✐...

1 ViT

Read the most influential vision transformer paper "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" (`https://arxiv.org/abs/2010.11929v2`).

2 DETR

Transformer-like architecture can be applied not only in the image classification task, but also in the fine-grained object detection task. Read DETR paper titled "End-to-End Object Detection with Transformers" (`https://arxiv.org/abs/2005.12872`).

Question:

Explain how transformers could be used for object detection. Can you briefly discuss what the object queries are in the DETR's decoder?