BABEŞ-BOLYAI UNIVERSITY CLUJ-NAPOCA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
SPECIALIZATION APPLIED COMPUTATIONAL INTELLIGENCE

DISSERTATION THESIS

Anchorless object detection based on deep convolutional networks

Supervisor
Prof. dr. Dioșan Laura

Author
Comănac Dragoș-Mihail

2024

ABSTRACT

Context

The purpose of artificial intelligence is to replicate the human intelligence, but with the advent of Large Language Models, the focus has shifted towards Natural Language Processing (NLP). We need to keep in mind that language is a high level abstraction of the real world created by the human mind, which is most likely not enough to build an Artificial General Intelligence (AGI) that can fully replicate the human intelligence.

One reason for this would be that language has a very limited bandwidth in terms of data. The human optical nerve is able to transmit to the brain data on a bandwidth of about 1 megabyte per second, which in one year would be approximately 30 terrabytes of data. This means that a small kid ”sees” more data than most LLMs ”see” during training, and this is without counting the other senses. Thus we believe that vision plays a crucial role in our understanding of the real-world, therefore it should play an important role in building the first AGI.

One way to make machines understand visual inputs such as images or videos is through computer vision. There are many tasks here such as image classification or segmentation, but for our work we will focus on object detection.

Therefore, we aim to build a framework based on Tensorflow and Keras which contains the necessary tools for building more complex applications of neural networks such as object detection. Also, we aim to implement in this framework one of the most recent approaches, namely anchorless object detection. In order to do this, we will take inspiration from one of the first approaches to do this: Fully Convolutional One-Stage Object Detection (FCOS). Here we also propose an optimization in the loss computation which we name ClipLoss, consisting of clipping the regression outputs of the detector to the shape of the image.

Regarding the dataset, we mainly use the 2D object detection subset of the KITTI dataset, which contains data specific to autonomous driving tasks, on which we achieve 28.7% mAP50.