Final Report

Ziyue Zeng[a]

[a]Waseda University, Shinjuku City, Tokyo, Japan

Abstract

This project is proposed by a team of Haoyang Tao and Ziyue Zeng. This proposal outlines a plan to implement the You Only Look Once version 1 (YOLOv1) object detection algorithm and train it on Tsinghua-Tencent 100K (Traffic-Sign Detection and Classification in the Wild) dataset. The goal is to achieve high accuracy in detecting objects of interest. After training, we will compare the performance of YOLOv1 with that of YOLOv6, a more recent version of the algorithm. This comparison will allow us to evaluate the improvements made in the YOLO algorithm over time and determine which version is best suited for our needs.

Keywords: Computer Vision, Object Detection, Image Classification

1. Background

Object detection is a crucial task in computer vision, with applications in various fields such as autonomous driving, surveillance, and image search. The YOLO (You Only Look Once) algorithm is a popular object detection method that has undergone several improvements since its first version, YOLOv1. This research aims to implement YOLOv1 and train it on the Tsinghua-Tencent 100K dataset to achieve high accuracy in detecting objects of interest. The performance of YOLOv1 will be compared with that of YOLOv6, a more recent version of the algorithm, to evaluate the improvements made over time and determine which version is best suited for the researchers' needs.

2. Objectives

This is the baseline of the project. However, it is possible for us to slightly modify or add new items into the objective during the process. If we made any change, they will be reflected in our final report.

1. To implement the YOLOv1 object detection algorithm and train it on the Tsinghua-Tencent 100K dataset.
2. To achieve high accuracy in detecting objects of interest using YOLOv1.
3. To compare the performance of YOLOv1 with that of YOLOv6, a more recent version of the algorithm.
4. To evaluate the improvements made in the YOLO algorithm over time and determine which version is best suited for the researchers' needs.

3. Methodology and Planning

3.1. YOLOv1 reproduction

We will start to learn the algorithm of yolo series from v1. As a real-time fast target recognition algorithm, the accuracy of yolov1 is not high enough. We applied it to the Tsinghua-Tencent 100K dataset this time, which has 221 categories. There are only 20 classification categories in the original work, and we will adjust and improve some parameters of the model. At the same time, some training techniques will also be used to improve the recognition accuracy.

3.2. Follow-up study

Afterwards, we will learn the improvement of each version of the yolo series algorithm to understand the improvement ideas and focus of the image recognition algorithm. This will allow us to gain experience and apply these techniques in our later research.

3.3. Comparison

Using the same data set in yolov6, the difference between the recognition results of the v6 version and the v1 version should be summarized.

3.4. Video recognition application

Finally, we use a yolo series algorithm to recognize a small video, taking advantage of its fast computing speed and real-time characteristics.

3.5. pre-implementation

Try to implement a real-time target recognition project combining camera and yolo algorithm. This can highlight the real-time characteristics of the yolo algorithm.