Digital Image Processing and Analysis Project Report:

Object Detection using YOLO and Faster R-CNN

Lovro Akmačić

0036537589

Faculty of Electrical Engineering and Computing

lovro.akmacic@fer.hr

Filip Buljan

0036539840

Faculty of Electrical Engineering and Computing

filip.buljan@fer.hr

Lucia Crvelin

0036540219

Faculty of Electrical Engineering and Computing

lucia.crvelin@fer.hr

Tomislav Čupić

00365411259

Faculty of Electrical Engineering and Computing

tomislav.cupic@fer.hr

Lorena Švenjak

0119039893

Faculty of Electrical Engineering and Computing

lorena.svenjak@fer.hr

May 29, 2025

Abstract

This report details a digital image processing and analysis project focused on object detection. We explore and implement state-of-the-art deep learning models, specifically YOLO (You Only Look Once) and Faster R-CNN (Region-based Convolutional Neural Network), for their efficacy in accurately identifying and localizing objects within images. The project aims to compare the performance, advantages, and limitations of these two prominent architectures in a practical application. We will discuss the theoretical underpinnings of each method, their practical implementation, and the insights gained from their application to a relevant dataset.

1 Description of Area and Activity

Digital image processing and analysis is a multidisciplinary field at the intersection of computer science, engineering, and mathematics, focusing on methods for manipulating and analyzing digital images. Its applications are vast, ranging from medical imaging and remote sensing to autonomous vehicles and security systems.

Our project delves into the sub-area of object detection, a fundamental task in computer vision that involves identifying instances of semantic objects of a certain class (e.g., humans, cars, animals) in digital images or videos and localizing them by drawing bounding boxes around them. The primary activity of this project involves the application and comparative analysis of two prominent deep learning-based object detection frameworks: YOLO and Faster R-CNN.

2 Overview of Methods Used

This section will provide a detailed technical overview of the two primary object detection methodologies employed in this project: YOLO and Faster R-CNN.

2.1 YOLO (You Only Look Once)

YOLO is a single-shot detector known for its speed and real-time processing capabilities. Unlike traditional object detection systems that separate the object detection pipeline into distinct stages (e.g., region proposal, classification, non-maximum suppression), YOLO processes the entire image in a single pass. It divides the image into a grid and simultaneously predicts bounding boxes and class probabilities for each grid cell. We will discuss its architectural components, including the backbone network, detection heads, and the loss function used for training. Different versions of YOLO (e.g., YOLOv3, YOLOv4, YOLOv5, YOLOv7) will be briefly mentioned, highlighting their key improvements.

2.2 Faster R-CNN (Region-based Convolutional Neural Network)

Faster R-CNN is a two-stage object detector that significantly improved upon its predecessors (R-CNN and Fast R-CNN) by introducing the Region Proposal Network (RPN). The RPN is a fully convolutional network that simultaneously predicts object bounds and objectiveness scores at each position. The proposed regions are then fed into the Fast R-CNN detection network for classification and bounding box regression refinement.

2.2.1 Fine tuning Faster R-CNN

3 Results

This section presents the results of our experiments with YOLO and Faster R-CNN on a chosen dataset.

4 Conclusion

References