Real-Time Object Detection using an Ensemble of One Stage and Two Stage Object Detection Models with Dynamic Fine-tuning using Kullback-Leibler Divergence

Lenard Nkalubo¹, Rose Nakibuule², and Nixson Okila³

¹Kyambogo University
²Makerere University
³Lira University

June 27, 2023

Abstract

Real-time object detection is a very challenging task, as it requires both high accuracy and high speed. One-stage object detectors such as YOLO models are very fast but they are also less accurate than two-stage object detectors such as Faster R-CNN. However, Faster R-CNN is not as fast as the YOLO models. In this study, we propose an ensemble approach to real-time object detection that combines the strengths of YOLOv5 and Faster R-CNN. We first use YOLOv5 to quickly generate a set of object proposals. We then use Faster R-CNN to refine these proposals and produce more accurate object detection results. To further improve the accuracy of our object detection results, we propose a cascade refinement network that uses dynamic fine-tuning. The cascade refinement network uses Kullback-Leibler divergence to dynamically adjust the weights of the Faster R-CNN model based on the confidence scores of the YOLOv5 object proposals. We evaluated our proposed approach on the novel dataset collected in Uganda with other State-of-the-art approaches which include RetinaNet, Cascade R-CNN, Single-Shot MultiBox Detector (SSD), and Region-based Convolutional Neural Network (R-CNN). Experimental results revealed that the proposed ensemble model outperformed both base models with an average precision of 0.96, which is significantly higher than the average precision of 0.91 for YOLOv5 and 0.90 for Faster R-CNN. The ensemble model was also able to achieve real-time inference speeds, processing frames at a rate of 25 frames per second, the same speed achieved by YOLOv5, faster than the speed of 15 frames per second by Faster R-CNN. The results also revealed that the proposed ensemble model is comparable to other state-of-the-art object detection models. Our proposed approach can be used to improve the accuracy and speed of real-time object detection in a variety of applications.

Hosted file