Object Detection and Key point Detection Using YOLOv8

1. INTRODUCTION

Object detection and human pose estimation are fundamental tasks in computer vision with applications in surveillance systems, sports analytics, healthcare monitoring, and human activity recognition. Recent advances in deep learning have significantly improved the accuracy and efficiency of these tasks. In this project, an object detection and pose estimation system was implemented using YOLOv8, developed by Ultralytics. The primary objective of this assignment was to construct a custom dataset from video clips, train a YOLO-based object detection model, evaluate its performance using standard metrics, and apply pose estimation to analyze human movements. The project demonstrates a complete deep learning workflow, including dataset preparation, annotation, training, evaluation, and result visualization.

2. DATASET DESCRIPTION

The dataset used in this project was created from six video clips containing **weight lifting activities**. From these videos, a total of 66 frames were extracted for annotation. Each frame was manually labeled using LabelImg software to labeled bounding boxes in YOLO annotation format, where each label file includes the class ID along with normalized values for the bounding box center coordinates, width, and height. The dataset consists of a single object class, namely "person." The images were divided into training and validation sets using an approximate 80:20 ratio to enable proper model evaluation.

A YAML configuration file was created to define the dataset structure, including the paths to the training and validation image directories, the number of classes (nc = 1), and the class name. The dataset link is provided separately as required in the submission guidelines. Although the dataset is relatively small, it is sufficient to demonstrate the implementation and evaluation of the object detection pipeline.

3. METHODOLOGY

3.1 Object Detection Model

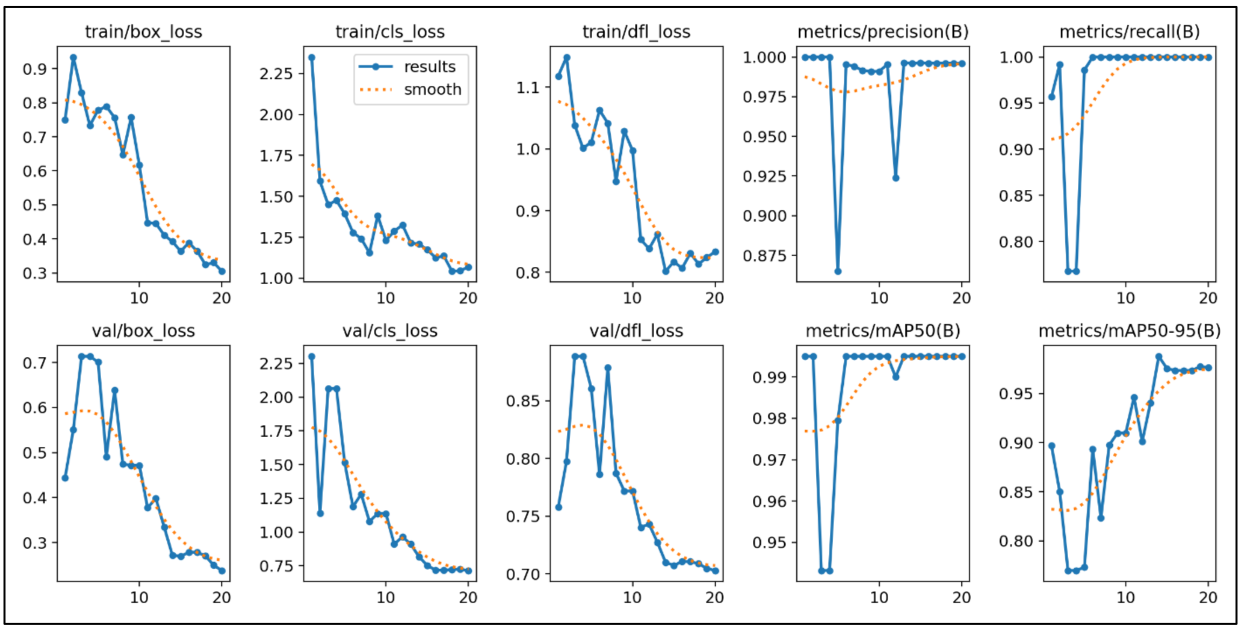
The object detection model was implemented using the YOLOv8 nano (YOLOv8n) architecture due to its lightweight design and suitability for training on small datasets. The model was trained for 20 epochs with an input image size of 640 × 640 pixels. During training, the framework automatically selected the AdamW optimizer and optimized hyperparameters. The training process monitored multiple loss components, including box loss for bounding box regression, classification loss for predicting the correct class, and distribution focal loss (DFL) for improved localization accuracy.

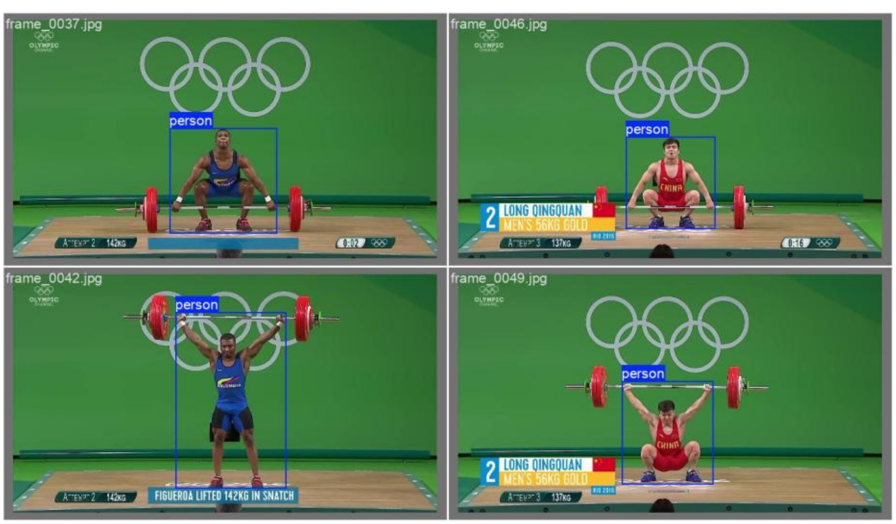
The training results included performance metrics and loss curves that illustrate the learning progression across epochs. The decreasing trend in box loss and classification loss indicates that the model successfully learned to localize and classify the object within the images. Sample detection outputs were generated to visually verify bounding box accuracy.

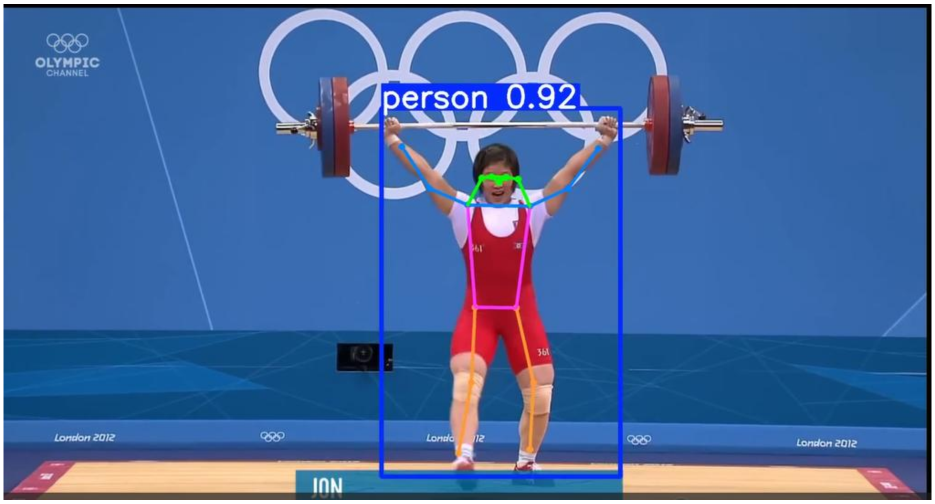
3.2 Pose Estimation

Following object detection, pose estimation was performed using a pretrained YOLOv8 pose model. This model detects multiple human keypoints, including shoulders, elbows, wrists, hips, knees, and ankles. The pose model was applied to all six video clips, generating output videos with skeleton overlays that represent detected keypoints and their connections.

The results demonstrate consistent tracking of human joints throughout the weight lifting movements. The pose estimation outputs effectively capture body posture and motion, showing the capability of the model to analyze human activity in dynamic video sequences.







4. CONCLUSION

In conclusion, this project successfully implemented a complete deep learning pipeline for object detection and pose estimation using YOLOv8. The custom dataset was created and annotated manually, and the detection model was trained and evaluated using standard performance metrics. The results demonstrate satisfactory detection accuracy and stable pose tracking in video sequences. Despite limitations related to dataset size, the project effectively illustrates practical understanding and application of modern object detection and pose estimation techniques in real-world scenarios.