M.Sc. in Computer Science

Postgraduate Institute of Science | University of Peradeniya

Semester II -2024/25

SC 549: Neural Networks

Programming Assignment – 03

S. Puvikanth

CSC 2417

Datasets and Resources: https://github.com/spuvi86/SC549-Programming-Assignment-03-CSC2417

1. Dataset Description

The dataset used for this assignment consists of video clips from three distinct sports: Cricket, Football, and Rugby.

- **Source**: Video clips provided in the video_clips directory.
- **Structure**: Frames were extracted from these videos to create a validation set (datasets/sports/images/val).
- **Classes**: The model is configured to detect a single class: person.

2. Methodology

The YOLO used for this as a real-time object detection from video clips.

- Models Used:
    - **yolo11n.pt (Nano):** Used for initial person detection.
    - **yolo11n-pose.pt (Nano Pose):** Used for keypoint estimation to understand player poses.

3. Frame Processing

The video processing pipeline as follows:

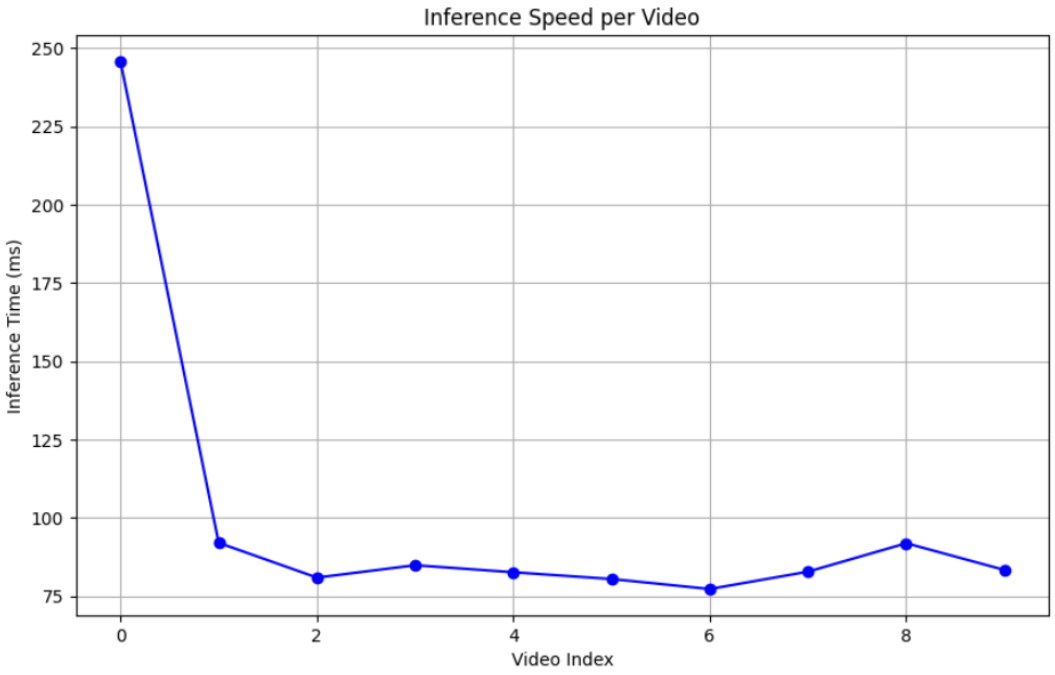
1. **Input**: Video files are read frame-by-frame using OpenCV.
2. **Inference**: Each frame is passed to the YOLO pose estimator.
    - *Resize*: Frames are automatically resized to the model's input dimension (640x640) by the ultralytics library.
    - *Prediction*: The model predicts bounding boxes, confidence scores, and keypoints.
3. **Drawing**: Annotated results (boxes and skeletons) are drawn onto the frames.
4. **Output**: Processed frames are saved as images (for the first few seconds) or compiled into a video (optional).

4. Model Performance and Analysis

A. Evaluation Metrics

Metric	Value	Description
mAP50-95	0.946	High consistency between detections and pseudo-labels.
Recall	0.943	The model successfully retrieves almost all instances.

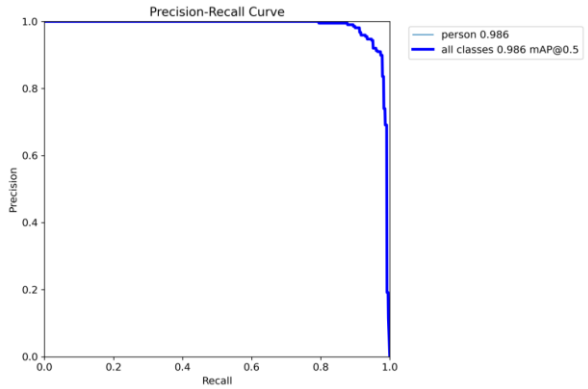
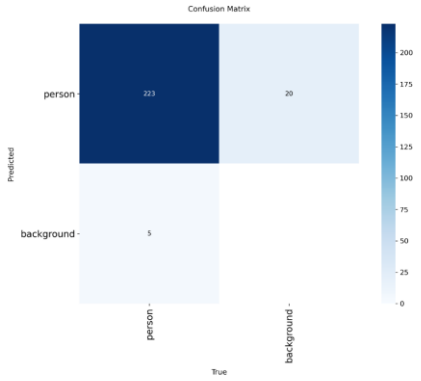
B. Inference Speed & Consistency



Inference Speed

Video Category	Avg. Inference (ms)	Observations
Cricket	~110 ms	Moderate speed, fewer players allows for high precision.
Football	~75 ms	Faster processing, varies with crowd density.
Rugby	~80 ms	Consistent speed despite heavy occlusion and dynamic movement.

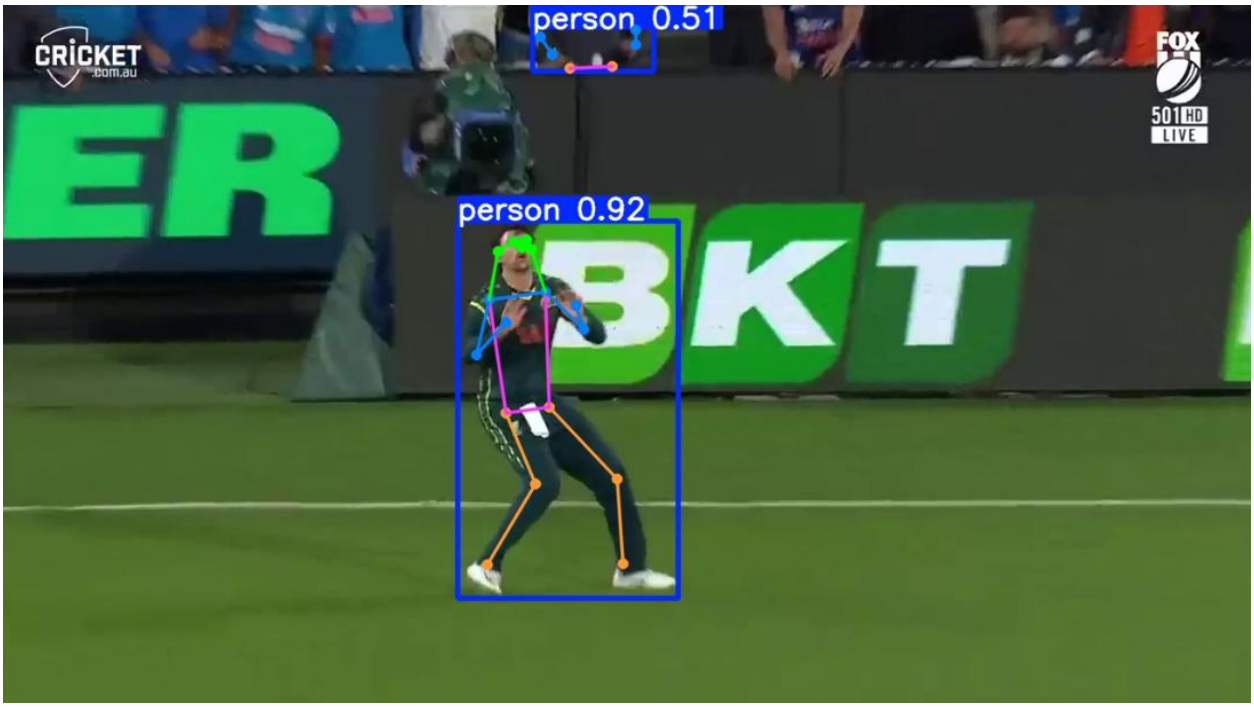
C. Visual Performance (Confusion Matrix & Curves)

Precision-Recall Analysis	Confusion Matrix
	
Pose PR Curve	*Confusion Matrix*

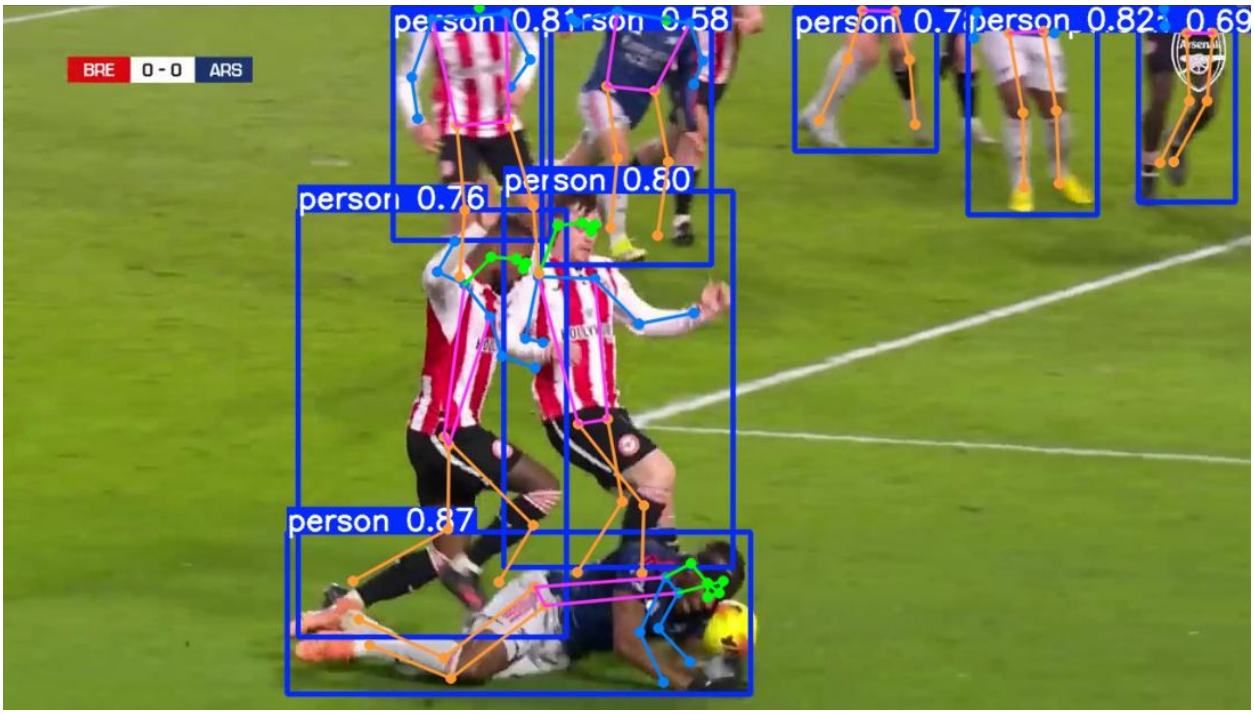
5. Output Images (Sample Results)

The following images demonstrate the model's detection capabilities across different sports.

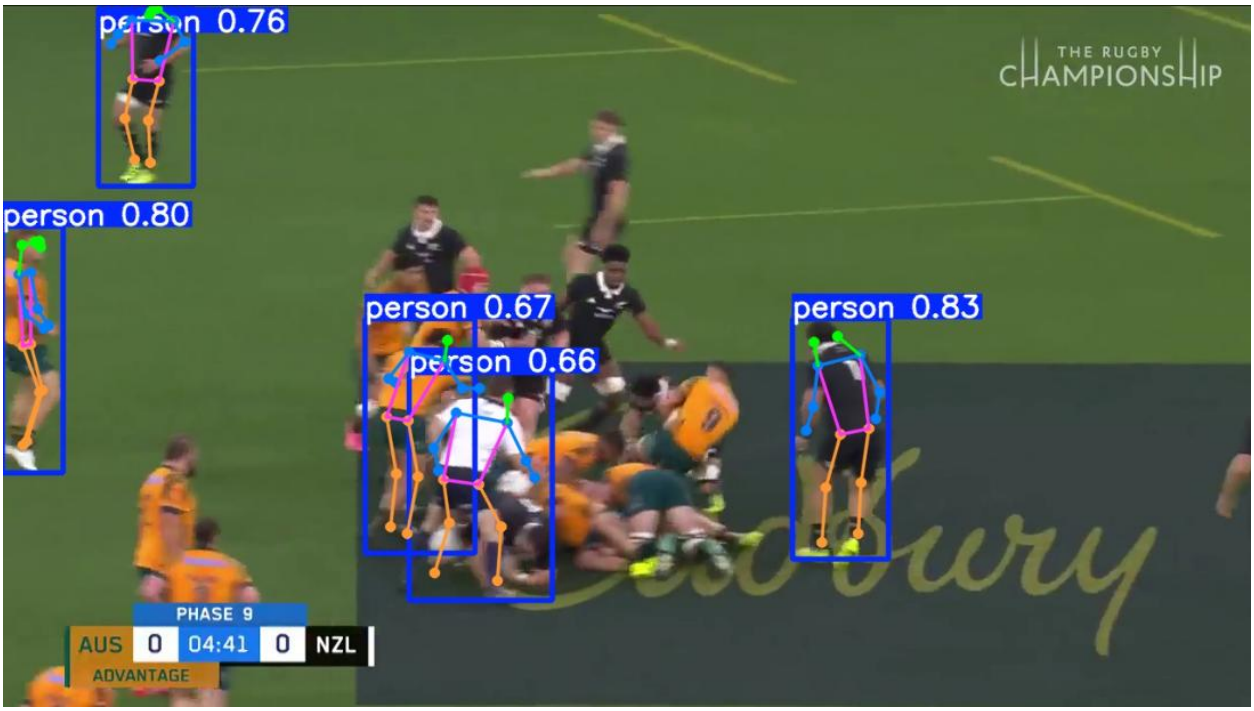
Cricket (*Detection of players on the field.)*



Football (*Multiple players detected in a crowded scene)*



Rugby (*Detection of players in dynamic poses)*



6. Possible Improvements

1. **Manual Labeling**: Replace pseudo-labels with manually verified ground truth to obtain scientifically accurate accuracy metrics.
2. **Model Size**: Upgrade to yolo11s (Small) or yolo11m (Medium) for better accuracy on small/distant players, at the cost of some inference speed.
3. **Tracking**: Implement object tracking to maintain player IDs across frames, which is crucial for sports analytics.
4. **Custom Training**: Fine-tune the model specifically on sports datasets (COCO-Keypoints or crowd-pose datasets) to better handle occlusion in rugby and football.