Basketball 3-Point Shot Detection & Pose Estimation Using YOLOv8 & MediaPipe

CSC2450_P P D A Dias

1. Introduction

Automated sports analytics has become an important application of computer vision and deep learning. In basketball, analyzing shooting mechanics and detecting successful 3-point shot attempts can provide valuable insights into player performance. This project focuses on detecting basketball 3-point shooting scenarios using object detection and human pose estimation techniques.

The system integrates YOLOv8 for object detection (player, ball and hoop) and MediaPipe Pose for extracting skeletal key points of the detected player. The objective is to build a complete pipeline capable of identifying shooting scenes and extracting biomechanical information from video frames.

2. Dataset Description

The dataset consists of 15 basketball videos focused on 3-point shooting scenarios. Frames were extracted from these videos and manually annotated using LabelImg. Annotations were created in YOLO format and the dataset was split into training and validation sets using an 80:20 ratio.

Dataset Summary:

- Total labeled images : **212**
- Training/Validation images : **171 / 41**
- Total object instances in validation set : **160**
- Classes : **Player, Ball, Hoop**

3. Methodology

The system consists of two main components:

1. Object Detection using YOLOv8
2. Pose Estimation using MediaPipe Pose

The pipeline follows the following structure.
Video → Frame Extraction → YOLO Detection → Player Cropping → Pose Estimation → Output Visualization

4. YOLOv8 Detection

YOLOv8n (nano) from the Ultralytics framework was used for object detection. The model was fine-tuned on the custom basketball dataset. After training, the best model weights were used for validation and inference.

Model Configuration:

- Epochs: 50
- Image size: 640 × 640
- Batch size: 8
- Optimizer: AdamW
- Hardware: NVIDIA Tesla T4 GPU (CUDA enabled)

5. Pose Estimation

MediaPipe Pose (33 keypoint model) was used to extract skeletal landmarks from detected player regions. This enables biomechanical analysis of shooting posture and provides structured spatial information for further analysis. The model follows the below procedure.

1. YOLO detects the player bounding box.
2. The player region is cropped.

3. MediaPipe Pose extracts 33 body landmarks.
4. Skeleton is drawn and saved.

6. Results & Performance Metrics

Validation results on 41 images (160 instances) are summarized below.

Metric	Precision	Recall	mAP50	mAP50-95
Value	0.914	0.862	0.929	0.524

The model achieved high detection accuracy at IoU = 0.5 (mAP50 = 0.929), indicating strong object localization and classification performance. Performance slightly decreased under stricter IoU thresholds (mAP50-95 = 0.524), which is expected due to tighter bounding box evaluation.

7. Loss Curve Analysis & Discussion

The training and validation loss curves show steady convergence. Precision and recall improved steadily across epochs, reaching approximately 0.91 and 0.86 respectively. mAP50 increased rapidly in early epochs and stabilized around 0.93, indicating good model convergence without severe overfitting.

- Box loss decreased consistently from approximately 1.9 to 1.0.
- Classification loss decreased from above 3.0 to below 0.8.
- DFL loss also showed a smooth downward trend.
- Validation losses followed a stable decreasing pattern with minor fluctuations.
- Player detection was highly accurate.
- Ball detection performed well but was occasionally missed due to motion blur and small object size.
- Hoop detection was consistent.
- Very few false positives were observed.
- GPU acceleration significantly improved training efficiency.

8. Limitations

Despite strong performance, several limitations exist.

1. Small dataset size (212 images) limits generalization.
2. Ball detection is sensitive to motion blur and scale variation.
3. Lighting and camera angle variations were limited.
4. Temporal information was not used (frame-by-frame detection only).
5. Pose estimation was applied only after detection, not temporally tracked.

9. Future Improvements

The system can be improved by increasing dataset size and diversity with data augmentation techniques. Object tracking and automated shot success detection can also be implemented. These enhancements would improve robustness and real-game applicability.

10. Conclusion

This project successfully implemented a complete computer vision pipeline for basketball 3-point shot analysis. YOLOv8 achieved high detection accuracy (mAP50 = 0.929) while MediaPipe Pose enabled detailed skeletal extraction of detected players.

The system demonstrates the effectiveness of combining deep learning-based object detection with pose estimation for sports analytics. With further dataset expansion and temporal modeling, this approach can be extended toward automated performance evaluation and real-time game analysis.