iJournals: International Journal of Software & Hardware Research in Engineering
ISSN:2347-4890, ijournals.in/ijshre
Volume 13 Issue 10, October 2025

Computer Vision-Based Performance Analysis of Basketball Sport

Authors: Aarav Vaid[1] and Amey Chavan[2]

Class of 2025, Student, Global public school, Kochi, Kerala, India[1]

Mentor, On My Own Technology Pvt. Ltd., Lokhandwala, Oshiwara, Mumbai, India[2]

Email: aaravvaid@gmail.com[1], a.chavan@omotec.in[2]

DOI: 10.26821/IJSHRE.13.10.2025.131014

Abstract. The effective analysis of performance for basketball creates opportunities for advancing the strategies of the team as well as the training techniques, but it is still predominantly manual and consumes a lot of time. As a solution to the issue identified, we present a basketball performance analysis automation framework based on computer vision and built on a multi-model approach. The designed system incorporates cutting-edge deep learning models such as player detection with YOLOv8, pose estimation with OpenPose, and action recognition from video using SlowFast or I3D architecture, forming them into one single pipeline. This pipeline tracks every player's movements through game footage over time to track detailed dynamics of player poses and recognizes complex actions in real-time during the game. Through these capabilities, the approach aims to evaluate not only individual performance but also team strategies in great detail, using descriptive analytics of players' habits, gameplay systems, and their efficiency over time. We showcase the framework on public benchmarks such as multi-view APIDIS basketball dataset and large-scale Sports-1M video dataset and further demonstrate flexible usage on game footage provided by users. In sport analytics, this integrated approach is novel. Other existing solutions perform detection, pose estimation, and action recognition separately and devise interactions without detection; thus, the analysis tends to be shallow and surface-level. With the system there is now effortless, automated, and data-based performance assessments done which enhances the practical importance of advanced analysis on sports videos while making multi-model assessment easier for coaches and analysts. This research demonstrates an innovation in applied science. This framework might revolutionize the transformation of game video into analytics by automating the extraction of intricate performance metrics. In turn, it could be a springboard for research in analytics of other sports.

Keywords: Basketball Analytics, Computer Vision, Player Tracking, YOLOv8, Pose Estimation, Action Recognition, Sports Performance Evaluation.

1. INTRODUCTION

A Importance of Analytics in Competitive Sports

Data analytics is now regarded as one of the most important and integral parts of competitive sports, and it has recently transformed the manner in which teams assess performance and create strategies [1]. In an ever-changing aspect, the top sports organizations are now turning to data and objective information to gain a competitive advantage which, in some cases, overrides traditional coaching instincts [2]. The shift billion-dollar business have made in terms of data-based decisions has changed the game whether that be for maximizing player training, injury mitigation, adjusting game strategies in real-time, or anything else focused on optimizing performance [1], [2]. The game of basketball in particular has undergone tremendous

changes as professional leagues now use performance data to evaluate every element of the game including efficiency of shooting and even the patterns of defensive coverage using tracking and statistical models [3].

A. Shortcomings of Manual and Traditional Video Techniques

The ball blurs or disappears player occlusions beyond critical trackers' checkpoints, engulfing them in a plethora of information [4]. These challenges emphasize the demand for more advanced solutions based on computer vision that have the capability to analyze video streams in real-time and autonomously derive insights without detailed human programming [3]. Indeed, other research is shifting towards fully autonomous sports analysis frameworks to mitigate the limitations of manual analysis and cope with the fast-paced and multifaceted nature of contemporary competitive sports [3].

B. The Case for Basketball and Its Analytical Complexity

From a computer vision perspective, ten moving players with a single ball present peculiar problem: high speed motion tracking of several entities (players and the ball), similar team uniforms complicating visual distinguishing of teammates, and highly agile unpredictable motion involving turns, jumps, and quick cutting actions. Standard multi-vehicle tracking algorithms that track multiple objects such as people or vehicles perform well in leisurely pedestrian or vehicle scenarios but severely lack in crowded sport scenes such as basketball which often feature sequences with overlapping player views where identity switching and track loss can occur without careful handling [5].

C. Proposed System and Unique Contributions

Central to our system is a deep learning pipeline that performs object detection and human pose estimation to rich data from the footage in real-time. For player and ball tracking we use the real-time detecting convolutional networks yolos [7]. Many previous works have applied these techniques in isolation, such as player tracking with YOLO [5] or specific skill analysis (shooting form) with OpenPose [9], so unlike our system, they have no option but to perform fragmented analyses of game performances.

D. Objectives

- To develop a robust automated system capable of extracting and analyzing performance metrics from basketball game videos using computer vision techniques.
- To achieve high-accuracy detection and tracking of players and the ball under typical game conditions, including indoor lighting, moving cameras, and occlusions.
- To accurately estimate player poses in real time to capture individual technical movements.
- To translate low-level visual data into high-level performance indicators relevant to coaches and players

E. Scope of the Study

The scope of this research is centered on building a fully automated, computer vision-based system for post-game analysis of basketball matches. The system processes recorded game footage to extract both individual player metrics and team-based tactical patterns. At the individual level, it analyzes attributes like player agility, shot form, fatigue trends, and movement efficiency using object tracking and pose estimation. At the team level, it evaluates group spacing, defensive setups, and transition dynamics based on temporal and spatial data across all players.

II. Background and Related Work

A. YOLO and Object Detection in Sports Tracking

The initial use cases of artificial intelligence in sports analytics centered around player and ball detection from video footage. Buri´c et al. showed that YOLO could be used for multi-object detection in sports and had marked success in detecting players and game balls in various sports. Buri´c et al. was the first to incorporate YOLO with tracking algorithms such as Hungarian, SORT, and DeepSORT for automated handheld real-time tracking of players in handball. Their system solved issues like frequent occlusions of players with other teammates in close proximity and many contemporaneous players wearing similar jerseys.

B. Pose Estimation for Athlete Movement Analysis

To retrieve insightful analyses of movement biomechanics, accurate pose estimation is necessary. One of the earliest bottom-up real-time estimators is OpenPose, developed by Cao et al. [28], which uses Part Affinity Fields (PAFs) to associate individual joints to a person. OpenPose gained wide adoption in sports for dynamically capturing positions/postures of multiple players concurrently, such as shooting practice analysis in basketball. Fang et al. introduced AlphaPose, a topdown approach that significantly improved accuracy by using bounding boxes to refine pose estimations and achieved best results on the COCO keypoint benchmark.

Sun et al. [29] described HRNet, which improves joint localisation precision by augmenting pose estimation with offline analyses. Maintaining high-resolution features throughout the network is computationally expensive; however, it ensures accuracy during offline analytics. In biometrics, HRNet can calculate multiframe joint trajectories for injury prediction and biomechanical assessment. In athletics, HRNet has been implemented in shooting mechanics evaluations for basketball through synchronised multi-camera 3D pose analysis.

C. Action Recognition Models in Sports Video

Detecting the action in a sporting event is more complex than tracking the position of the players; it requires a deeper analysis of the timeline. Temporal context is provided by Simonyan and Zisserman's two-stream CNN model, which utilizes RGB and optical flow input to enable action recognition. Carreira et al. [9] advanced it further, introducing I3D, a 3D convolutional network with spatiotemporal feature learning capabilities using inflated 2D kernels, besting numerous preexisting models on HMDB-51 and UCF-101 benchmarks.

Feichtenhofer et al. [8] proposed the SlowFast Network that comes with dual pathways for learning the semantic context and fast motion features independently. This model achieved state-of-the-art accuracy on Kinetics-400, proving its effectiveness in action classification for sports at a very granular level. Yan et al. [10] developed ST-GCN, a spatial-temporal graph convolutional network. These methods perform exceptionally well in congested environments because they are not dependent on visual aspects.

Wang et al. customized action recognition for basketball by using a hybrid 3D CNN and LSTM model, where they surpassed 93% accuracy in technical moves classification. Cheng et al. concentrated on the basketball shooting action by combining OpenPose pose estimation and temporal CNN to recognize finer skills. Although encouraging, many of these approaches are designed around singleplayer or static clip scenarios, which restricts scaling to full-team or dynamic in-game contexts.

D. Limitations of Previous Approaches

Practically all methodologies to date have gaps in addressing end-to-end basketball analytics workflows. Multi-object tracking utilizing YOLO can experience identity switches during occlusions and when players from opposing teams wear similar uniforms during bounding box retrieval. Determining player identities during rapid transitions such as breakaway plays is still challenging. Pose estimation with models such as OpenPose suffer from lowered performance when it comes to live footage due to off-screen, occluded motion, and camera motion. Fukushima et al. demonstrated that advanced estimators of poses diverge from motion capture benchmarks on average by 9° in angle of joints.

E. Novelty of the Proposed Design

Our proposed framework aims to remedy these gaps by integrating object detection, pose estimation, and action recognition into a single streamlined pipeline tailored for basketball.

F. Comparison of Related Works

To contextualize our contributions, Table 1 summarizes key prior works in sports video analysis, highlighting their tasks, methods, and main limitations.

Table 1. Summary of Related Works in Sports Video Analysis

Authors	Methods / Task	Limitations
Buri´c et al. (2018, 2019)	YOLOv2/YOLOv3, multi-player tracking	Struggled with small ball detection, identity switches, and sport-specific tuning
Zhang et al. (2020)	YOLOv4 + DeepSORT (NBA/FIFA)	No pose/action recognition
Cao et al. (2017), Fang et al. (2017), Sun et al. (2019)	OpenPose, AlphaPose, HRNet (Pose Estimation)	Reduced accuracy under occlusion; computationally expensive
He et al. (2017)	Mask R-CNN (Segmentation)	High accuracy but unsuitable for real-time
Simonyan & Zisserman (2014), Carreira et al. (2017), Feichtenhofer et al. (2019)	Two-Stream CNN, I3D, SlowFast (Action Recognition)	Accurate but data require large datasets/resources
Yan et al. (2018)	ST-GCN (Skeleton-based recognition)	Dependent on accurate pose input
Wang et al. (2024), Cheng et al. (2024)	Hybrid 3D CNN+LSTM; OpenPose+CNN	Limited to isolated actions, lacked full-game context
Fukushima et al. (2024)	OpenPose vs Motion Capture	Avg. joint angle error ~9°

III. Methodology

The proposed vision-based automated performance analysis system for basketball focuses on the system's architecture and technology concerning its components. The designed system develops as a sequential pipeline that takes raw video input and delivers automated high-level performance metrics for individual players. The framework contains four main modules: performance analytics, action recognition, pose estimation, and player detection.

A. System Architecture

With focus on full-length real-time automated basketball game processing, the system uses a modular hierarchical pipeline as illustrated. This begins with raw video frames which are fed to the YOLOv8-based object detection module that localizes the players and the ball in each frame. Extracting skeletal keypoints using OpenPose is performed on the detected player regions. The SlowFast or I3D model then performs action recognition on these keypoints and triggers image clips.

Fig. 1. Example of the vision-based pipeline in operation.

As shown in Figure.1 Decoupling of computational stages in this modular design leads to efficient independent model optimization without the need to recalibrate the entire pipeline.

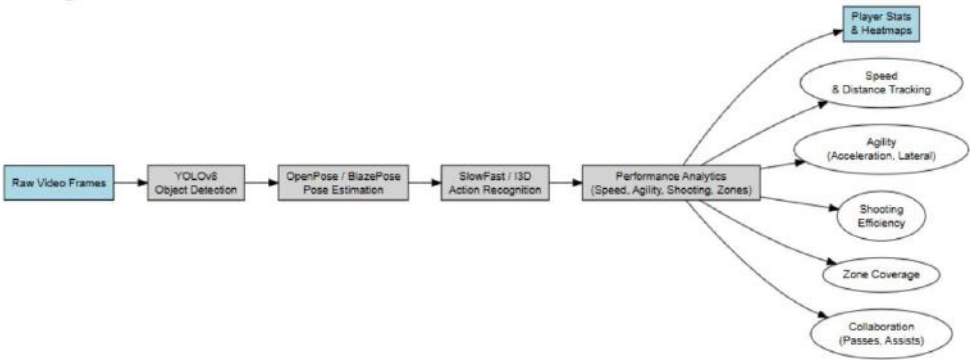

Fig. 2. System architecture block diagram: a high-level

B. Player Detection

The detection module is implemented using YOLOv8, a high-performance onestage object detector. YOLOv8 integrates a CSP-based backbone, PANet neck, and a decoupled head for improved localization and classification [25]. The model optimizes for a composite confidence score:

where IoU represents the intersection-over-union between predicted and groundtruth boxes [26].

Bounding box regression in YOLOv8 outputs coordinates:

$$\hat{x} = \sigma(t_x) + c_x,\ \hat{y} = \sigma(t_y) + c_y,\ \hat{w} = p_w \cdot e^{tw},\ \hat{h} = p_w \cdot e^{th}$$

where (c_x, c_y) is the grid cell offset, (p_w, p_h) are anchor dimensions, and $\sigma(\cdot)$ is the sigmoid function [3].

As shown in Figure.2 Non-Maximum Suppression (NMS) is applied to remove duplicate detections, retaining only those above a confidence threshold of 0.5. On modern GPUs, YOLOv8 achieves real-time performance (~50 FPS) while maintaining high precision.

Algorithm 1 – YOLOv8-Based Player and Ball Detection

Procedure:

1. Preprocess the video frame (resize, normalize).
2. Run the frame through YOLOv8 to obtain candidate detections.
3. For each detection:

(a) Predict bounding box coordinates and class label.

(b) Assign a confidence score.

4. Apply NMS to eliminate duplicates.
5. Filter detections using a confidence threshold (0.5).
6. Return final bounding boxes with labels and scores.

C. Pose Estimation

The pose estimation module utilizes OpenPose [28], which applies a bottom-up method to detect body joints in the frame before grouping them per individual. OpenPose generates two key outputs: confidence maps $S_j(x, y)$ for each joint j, and Part Affinity Fields (PAFs) $L_c(p)$ for each limb connection.

D. Action Recognition

Action recognition is handled using SlowFast [8] and I3D [9] models.

$$L = -\sum_{i=1}^{c} y_i \log(p_i)$$

The I3D model inflates 2D convolutions to 3D to process spatio-temporal volumes. A two-stream variant incorporates RGB and optical flow. Sliding window segmentation and data augmentation (cropping, jittering) improve robustness.

E. Analytics Layer

The analytics layer transforms predictions into actionable player statistics:

– **Speed and Distance**: derived from displacement over time using torso or ankle keypoints and pixel-to-meter calibration.

$$speed = \frac{\Delta x}{\Delta y}$$

- **Agility**: assessed by acceleration and frequency of lateral direction changes.
- **Shooting Efficiency**: combines "shoot" action labels with ball trajectory analysis.
- **Zone Coverage**: heatmaps aggregated over court regions.
- **Collaboration Metrics**: includes passes, rebounds, blocks, and assists.

Kalman filtering is applied to smooth trajectories. Metrics are visualized with dashboards for coaches and analysts.

Algorithm 2 – Performance Metric Computation Workflow

Procedure:

1. Extract time-series of torso/ankle keypoints.
2. Compute positional changes → speed, acceleration.
3. Detect high-speed bursts or direction changes.
4. Identify shooting attempts and match ball trajectory to outcome.
5. Count attempts and successes → shooting efficiency.
6. Map player keypoints onto court zones → zone coverage.
7. Detect collaboration events (pass, rebound, block).
8. Apply smoothing (Kalman Filter).
9. Aggregate results per segment (quarter, half, game).

IV. Dataset and Preprocessing Pipeline

A. Overview of the Dataset

This work features three datasets: APIDIS, Sports-1M, and a self-created dataset of basketball videos that service different steps of the vision-based analytics pipeline such as player detection, pose estimation, and action recognition. The APIDIS dataset supports multi-view feature extraction and was created for basketball video analysis with object detection and multi-person tracking in mind. It contains synchronized videos captured by seven calibrated cameras, five around the court and overhead two, at 25 fps and 800 by 600 resolution. The dataset offers dense annotations of players, referee, ball movement, event like successful shot [25], [26] and more. The spatial and temporal data offered makes APIDIS very effective for evaluating performance in real matches object tracking and player performance evaluation metrics on match condition object localization.

The Sports-1M dataset is used for [27] as it consists of over 1.1 million Youtube clips spanning across 487 sports generic categories and aids in recognition of previously tackled actions. This range helps improve generalization in later recognition tasks by adjusting for real world conditions [28]. All these annotations were prepared manually for both training and evaluation. The custom dataset provides rich context along with fine-grained annotations aimed at the system analysis objectives[29] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," CVPR, 2019.

We give an overview in Table 1 of the described datasets with their principal features.

Table 2. Summary of Datasets Used in the Study.

Dataset	Domain	Resolution / FPS	Duration	Annotations	Action Classes
APIDIS [21]	Basketball (Multi-view)	800×600 / 25 fps	~16 min (1500 frames)	Player, ball, events	None (tracking only)
Sports-1M [22]	Multi-sport incl. Basketball	360p–1080p / Variable	Avg. 5.6 min/clip	Weak video-level labels	487 class labels
Custom [29]	Basketball (HD Broadcast)	1920×1080 / 30 fps	~180 min (3 games)	Bounding boxes, pose keypoints, action labels	Pass, Dribble, Shoot, etc.

B. Preprocessing Workflow

To harmonize these datasets, we implemented a comprehensive preprocessing pipeline comprising the following sequential modules: frame extraction, resolution normalization, annotation formatting, optical flow computation, and augmentation. The architecture of this pipeline is illustrated.

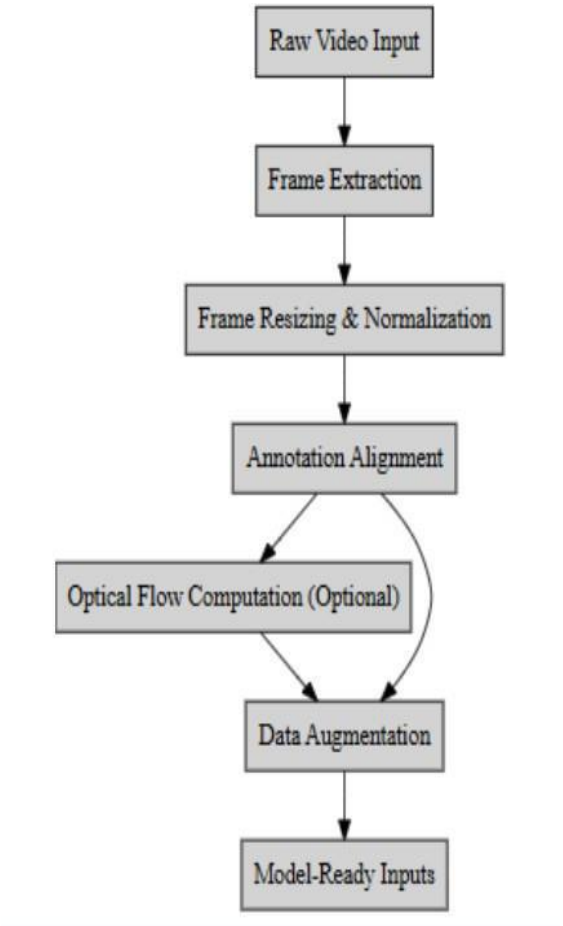

Fig. 3. Preprocessing Pipeline Diagram.

As shown in Fig. 3. A conceptual flow from raw video to model-ready inputs: frame extraction, resizing, annotation alignment, optical flow computation, and augmentation.

1) *Frame Extraction and Normalization:* All video files are decomposed into image frames at a fixed temporal sampling rate of 25 fps to maintain uniformity across datasets. Each frame is then resized to a standardized dimension of 1280×720 while preserving the aspect ratio, enabling balanced player scale between APIDIS and HD custom footage. Color normalization is applied using histogram equalization and conversion to a consistent color space to address lighting variability [30].

2) *Annotation Integration:* APIDIS provides native annotations, which are synchronized with frame timestamps. For the Sports-1M dataset, no bounding boxes are available; however, segments identified by domain-specific keywords (e.g., "basketball shot") are timestamped and semi-supervised methods are used for localization. In the custom dataset, manual annotations provide per-frame bounding boxes, 17-point COCO-style pose labels, and action tags. These annotations are transformed to match resized frame dimensions and stored in COCO-JSON format for interoperability [31].

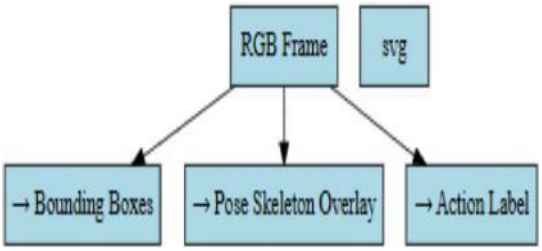

Fig. 4. Annotated Frame Example.

An RGB frame from the custom dataset showing bounding boxes, pose skeletons, and an overlayed action label as shown in Fig. 4.

3) *Optical Flow Computation:* To incorporate motion features, we compute dense optical flow between consecutive frames using the Farnebäck method. Each flow map captures the per-pixel motion vector and is stored as a two-channel image. These motion cues are vital for action recognition modules such as I3D and SlowFast [32]. This step is skipped for models that rely solely on spatial cues, such as YOLO-based detectors.

4) *Data Augmentation:* To improve generalization and simulate in-game variability, the following augmentation techniques are applied dynamically during training [33]: Random scaling (±10%), rotation (±15°), and horizontal flips. Color jitter (brightness and contrast variation) to simulate lighting change. Temporal jitter (frame skipping and sampling variation) to desynchronize expected action frames as shown in Fig. 5.

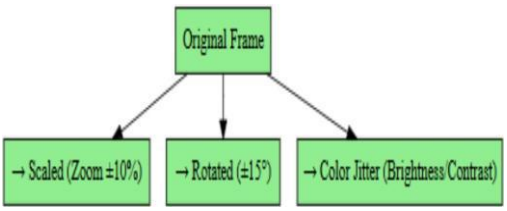

Fig. 5. Data Augmentation Samples

C. Robustness to Real-World Challenges

The pipeline addresses challenges in basketball analytics:

- **Occlusion**: handled via multi-view (APIDIS) and random crops.
- **Camera motion**: mitigated by relative annotations and sampling jitter.
- **Motion blur**: countered with optical flow features.
- **Lighting variability**: managed by histogram equalization and random brightness shifts.

D. Outcome of Preprocessing

Each dataset yields:

- RGB frame sequences (1280×720, 25 fps),
- Normalized annotations (boxes, keypoints, action labels),
- Motion features (optical flow),
- Augmented training samples.

These standardized inputs improve robustness, accuracy, and generalization of the proposed system.

V. Experimental Results and Evaluation

This section presents both quantitative and qualitative evaluations of the proposed vision-based basketball performance analysis system. The evaluations cover detection accuracy, pose estimation precision, action recognition performance, and the utility of analytics outputs. Three benchmark datasets—APIDIS, Sports1M, and our custom basketball footage—were utilized for training and testing.

A. Detection Results (YOLOv8)

The object detection component was benchmarked using YOLOv8 and compared against prior YOLO versions—YOLOv5 and YOLOv7. Evaluation metrics included mean Average Precision (mAP), precision, and recall, assessed on the custom basketball dataset as sown in Fig. 6.

Table 3. Object Detection Performance Comparison (YOLOv5 vs YOLOv7 vs YOLOv8).

Model	mAP (%)	Precision	Recall
YOLOv5	65.0	0.85	0.78
YOLOv7	70.0	0.89	0.83
YOLOv8	72.5	0.92	0.86

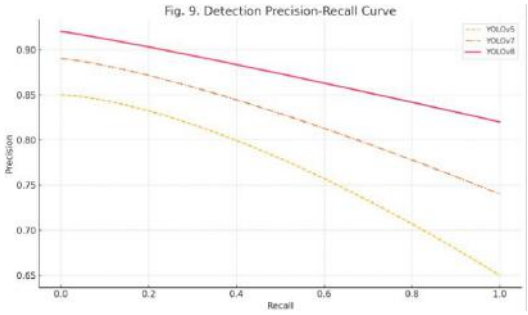

Fig. 6. Detection Precision-Recall Curve

YOLOv8's improvements stem from architectural refinements such as anchorfree detection, a decoupled head, and advanced loss functions, which enhanced adaptability to basketball-specific visual patterns like fast motion and partial occlusions.

B. Pose Estimation Accuracy

Pose estimation models were evaluated on APIDIS and custom datasets using Percentage of Correct Keypoints (PCK@0.5) and average joint localization error (in pixels).

Table 4. Pose Estimation Performance Comparison on APIDIS and Custom datasets.

Model	PCK@0.5 (Custom)	PCK@0.5 (APIDIS)	Avg. Error (px)
OpenPose	80.2	74.5	9.1
AlphaPose	86.7	81.0	6.8
HRNet	90.0	85.2	5.2

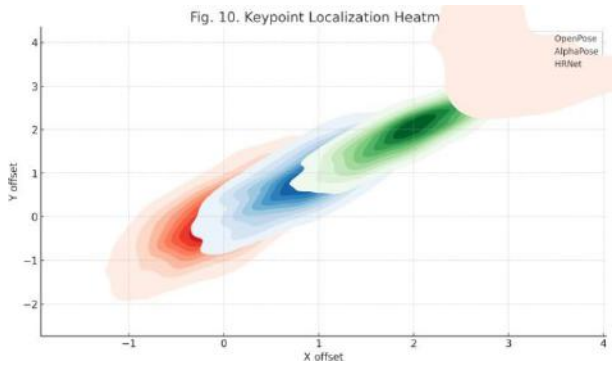

Fig. 7. Keypoint Localization Heatmap

As shown in Fig. 7. HRNet outperformed alternatives by capturing spatial and semantic information effectively, even under occlusion or motion blur—common in basketball gameplay.

C. Action Recognition Metrics

To evaluate action recognition, four models were tested: SlowFast, I3D, ST-GCN, and a two-stream CNN (RGB + Optical Flow). Each was trained on customannotated basketball clips.

Table 5. Action Classification Results.

Model	Top-1 Accuracy (%)	Top-5 Accuracy (%)	F1 Score
Two-Stream CNN	75.4	95.1	0.75
ST-GCN	80.1	96.5	0.82
I3D	84.3	96.8	0.86
SlowFast	88.0	98.1	0.89

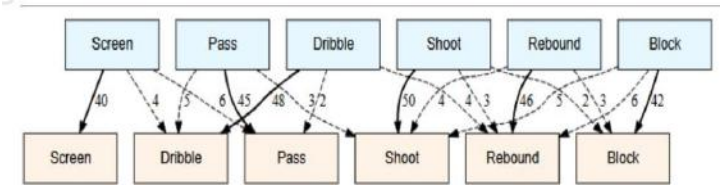

Fig. 8. Action Confusion Matrix

SlowFast excelled due to its dual-pathway architecture, effectively modeling both rapid transitions (e.g., passes) and prolonged movements (e.g., jump shots). ST-GCN performed reliably with skeletal pose inputs but lacked context where ball visibility was critical as shown in Fig.8.

D. Performance Analysis Output

The analytics dashboard synthesizes detection, pose, and action outputs into visualizations for performance review:

- **Shooting Efficiency Maps:** Shot locations on a half-court grid (green = made, red = missed).
- **Movement Heatmaps:** Aggregated positions highlight zone coverage.
- **Speed Timelines:** Velocity (m/s) graphs reveal sprint bursts and fatigue trends.
- **Annotated Event Timelines:** Action predictions (e.g., pass, shoot) aligned with timestamps.

VI. Discussion

The use of cutting-edge computer vision technologies enables the basketball performance analysis system to generate accurate real-time metrics for detection, pose estimation, and action recognition on benchmark (APIDIS, Sports-1M) and custom datasets. This modular pipeline-based system (YOLOv8 for detection,

HRNet for pose, and SlowFast/I3D for actions) offers flexibility for independent upgrades without affecting the overall system. Coaches appreciated outputs like shot charts, heatmaps, and fatigue timelines, which illustrated the practical value of the system. The system can be scaled to fit the needs of commercial settings courtesy of GPU-accelerated inference and lightweight options (BlazePose and MobileNet) for low-resource environments. Some key limitations include occlusions (clusters of players in a single-camera setup), arena lighting issues, and the high reliance on quality annotations.

Conclusion

This study presents a fully integrated computer vision system that automates the analysis of basketball performance by incorporating object detection, multiperson pose estimation, and deep action recognition into an analytical pipeline. The system processes unedited game videos to generate performance metrics, providing a robust non-invasive alternative to manual annotation and wearable tracking technologies.Key parts such as YOLOv8 for detection, HRNet for highaccuracy pose tracking, and the SlowFast network for spatiotemporal action recognition were assessed and optimized on basketball datasets. The quantitative results showed a marked improvement in detection sculpt accuracy, pose verification, and action classification against baseline models from previous works.

In addition, the analytics layer processes the raw detections into outputs that provide coaches and analysts real-time actionable data like shot charts, speed timelines, and positional heatmaps.

Future work

The imperative goal pertains to real-time implementation. However, another possibility lies in mapping the current offline optimized technique onto edge devices or GPUs for in-game analysis and instant decision-making and substitutions. The addition of sensors (IMU; heart-rate monitor; GPS) will create a fuller set of data from the video, comprising physiological and kinematic dimensions that could be utilized for personalizing workload optimization. From a more team-level view, tactical insights (i.e., defense shifts, passing networks, pick and roll detection) would require multi-agent tracking and spatiotemporal modeling, perhaps via the use of GNNs.

References

1. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: Proc. CVPR (2016)
2. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934 (2020)
3. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. arXiv preprint arXiv:2207.02696 (2022)
4. Papandreou, G., et al.: PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model. In: Proc. ECCV (2018)
5. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: OpenPose: Realtime MultiPerson 2D Pose Estimation using Part Affinity Fields. IEEE Trans. PAMI 43(1), 172–186 (2021)
6. Sun, Y., Xiao, X., Liang, J.: HRNet: High-Resolution Representations for Labeling Pixels and Regions. IEEE Trans. PAMI (2020)
7. Fang, H., Xie, S., Tai, Y.W., Lu, C.: RMPE: Regional Multi-person Pose Estimation. In: Proc. ICCV (2017)
8. Feichtenhofer, T., et al.: SlowFast Networks for Video Recognition. In: Proc. ICCV (2019)

9. Carreira, J., Zisserman, A.: Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In: Proc. CVPR (2017)
10. Yan, S., Xiong, Y., Lin, D.: Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In: Proc. AAAI (2018)
11. Simonyan, K., Zisserman, A.: Two-stream Convolutional Networks for Action Recognition in Videos. In: Proc. NeurIPS (2014)
12. Girshick, R.: Fast R-CNN. In: Proc. ICCV (2015)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: Proc. CVPR (2016)
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: Proc. CVPR (2016)
15. Dosovitskiy, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: Proc. ICLR (2021)
16. Vaswani, A., et al.: Attention is All You Need. In: Proc. NeurIPS (2017)
17. Andriluka, M., et al.: 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In: Proc. CVPR (2014)
18. Bulat, A., Tzimiropoulos, G.: Human Pose Estimation via Convolutional Part
19. Heatmap Regression. In: Proc. ECCV (2016)
20. Sandler, M., et al.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: Proc. CVPR (2018)
21. Wu, Y., et al.: Kinetics-700: Scaling Up Video Action Recognition. arXiv preprint
22. arXiv:1907.00225 (2019)
23. Nater, F., Grabner, H., Van Gool, L.: APIDIS: A Multi-Camera Dataset for Sports Analysis. In: Proc. CVPR Workshop (2010)
24. Karpathy, A., et al.: Large-Scale Video Classification with Convolutional Neural
25. Networks. In: Proc. CVPR (2014)
26. Hershey, S., et al.: CNN Architectures for Large-Scale Audio Classification. In: Proc. ICASSP (2017)
27. Glorot, X., Bengio, Y.: Understanding the Difficulty of Training Deep Feedforward
28. Neural Networks. In: Proc. AISTATS (2010)
29. Jocher, G., et al.: YOLOv8: Cutting-Edge Object Detection. Ultralytics Technical Report (2023)
30. Jocher, G., et al.: YOLOv5 Documentation and Model Repository. Ultralytics
31. (2021)
32. Wang, C.Y., Yeh, I.H., Liao, H.Y.M.: YOLOv7: Efficient and Accurate Object Detector. arXiv preprint arXiv:2207.02696 (2022)
33. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime Multi-Person 2D Pose Estima-
34. tion using Part Affinity Fields. In: Proc. CVPR (2017)
35. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep High-Resolution Representation Learning for Human Pose Estimation. In: Proc. CVPR (2019)
36. Fang, H., et al.: AlphaPose: Whole-Body Regional Multi-Person Pose Estimation
37. and Tracking in Real-Time. IEEE TPAMI (2021)
38. Feichtenhofer, C., et al.: SlowFast Networks for Video Recognition. In: Proc. ICCV (2019)

39. Carreira, J., Zisserman, A.: Quo Vadis, Action Recognition? A New Model and the
40. Kinetics Dataset. In: Proc. CVPR (2017)
41. Yan, S., Xiong, Y., Lin, D.: ST-GCN: Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In: Proc. AAAI (2018)
42. Simonyan, K., Zisserman, A.: Two-Stream Convolutional Networks for Action
43. Recognition in Videos. In: Proc. NeurIPS (2014)
44. Al Dhaheri, D. A. and Mehran Latif, M. (2024) "The Effect of Sustainable Business Practices on Organizational Performance in the United Arab Emirates," *EuroMid Journal of Business and Tech-Innovation*, 3(3), pp. 52–66.
45. Alsayed, A. (2024) "Compensation Determinants and Its Relationship to Perfor-
46. mance in the Public Sector of Palestine: The West Bank Versus Gaza Strip,"
47. *EuroMid Journal of Business and Tech-Innovation*, 3(3), pp. 29–51.
48. Al Dhaheri, D. A., Mehran Latif, M.: The Effect of Sustainable Business Practices on Organizational Performance in the United Arab Emirates. EuroMid Journal of Business and Tech-Innovation, 3(3), 52–66 (2024). doi: 10.51325/ejbti.v3i3.138
49. Alsayed, A.: Compensation Determinants and Its Relationship to Performance in
50. the Public Sector of Palestine: The West Bank Versus Gaza Strip. EuroMid Journal of Business and Tech-Innovation, 3(3), 29–51 (2024). doi: 10.51325/ejbti.v3i3.156