[1]Mingsheng Xu

# Basketball Tactics Analysis Based on Improved Openpose Algorithm and Its Application

Journal of Electrical Systems

***Abstract:*** - The use of information technology has currently resulted in a change in people's production and lifestyles as well as a catalyst for the sports industry's evolution. Currently, there is a growing use of digitization in the sports industry's basketball realm. Using 2D posture estimation, a novel lightweight deep learning architecture (LDLA) is built to accomplish the automatic analysis of basketball game footage. We offer a real-time method to identify numerous people's 2D poses in a video. Additionally, it examines how 2D pose estimation might be used to analyse basketball shooting videos. First, group and global motion features are extracted to represent semantic events. A full basketball game video is broken down into three stages: clip-based segmentation, classification of semantics occurrences utilizing audio and video, and temporal sequence characteristics employing GRU CNN. Video is first processed to create some basic video categorization and segmentation using the visual, movement, and auditory data. Further use of domain knowledge is made to find important events in the basketball video. Shot and image threshold recognition algorithms for videos have used both optical and kinematic prediction information; scene classification comes next. The placements of probable semantic events, such as "fouling" and "shooting at the basket," are then discovered by comparing the multidimensional data with supplementary domain expertise. Experimental results demonstrate the proposed LDLA method achieves 99.6% of accuracy, 93.2% of precision, 93.5% of recall and 89.5% of F1-score.

***Keywords:*** smart court, basketball, neural network, pose estimation, OpenPose, global motion.

## I. INTRODUCTION

Dribbling is the most fundamental activity in basketball, and shooting is the only way to win the game. Dribbling, shooting, and layups are all basic actions in basketball games [1]. The game's major impact score is greatly influenced by how well the basic actions are performed. Combining human pose estimation and action recognition algorithms has become increasingly important as basketball games have developed [2]. This helps to increase the scoring rate. Human pose estimate is the process of identifying and estimating each part of the target human body from the image in terms of its position, orientation, and scale[3]. This data must be transformed into a digital format so that the computer can understand it and output the present human posture and action. Action recognition is utilized as the input object to determine whether a person's behaviors are acceptable and how they can increase the normativeness, however it is reliant on the outcome of posture estimate [4]. In team sports, finding the ball is a crucial addition to player identification and tracking, both to feed sport analytics and to improve broadcast content. Knowing the position of the ball accurately and without delay is much more crucial in the context of real-time increased automation of team major sporting events [5]. The difficulty of finding the ball has been extensively studied over the past 2 decades, leading to commercial goods like the computerized handball line identifying technology or the goal-line equipment in baseball [6]. However, due to two key problems, the ball identification problem is still open for cases of significant practical significance. First of all, as is frequently the case in organized sports, the endeavor becomes quite difficult when the ball is partially obscured by player interactions [7]. A standardized multi-view collection configuration has typically been taken into account in previous works to tackle this problem. This arrangement increases the identification effectiveness of contenders by ensuring uniformity among perspectives. Several studies have indeed suggested enhancing multiview ball identification with cues generated from player tracking to better handle situations when the ball is handled by players [8]. These methods, nevertheless, call for the placement of numerous cameras all throughout the playing area, which is significantly more costly than single perspective collection systems and seems challenging to implement in many settings. Inertial device sign language [9] and picture gestures [10] are the two primary kinds of basketball gesture recognition techniques. The inertial camera pose cohesion necessitates the use of the sensor by the athlete, who then sends the data obtained to the data processing terminal for action recognition analysis.

[1] Department of physical education, Yangzhou Polytechnic Institute, Yangzhou, Jiangsu, China, 225127

*Corresponding author email: xumingsheng85@163.com

This is a significant amount of equipment and is not suitable for widespread deployment. The pose classification for picture acquisition starts by using video or images that have been taken by the camera, then extracts hidden characteristics from the video or images, and, ultimately, applies a classifier to perform motion identification. Given such, the work's contributions are just as follows:

- Using the Gated Recurrent Unit Convolutional Neural Network (GRU-CNN) and the clip-based segmentation method, the spatial properties of global motion patterns are retrieved.
- Researchers had integrated the audio and motion data with additional low-level characteristics like hue and texture to enable comprehensive semantic basketball visualization techniques and annotation.

The paper is organized as follows. Section 2 describes about some of the existing methods for constructing smart ground for sports with efficient video analysis. Section 3 shows the overall classification framework in basketball game videos, based on multidomain knowledge. Section 4 presents the experimental results that quantify the performance of the proposed approach. Finally, conclusions are drawn in Section 5.

## II. RELATED WORKS

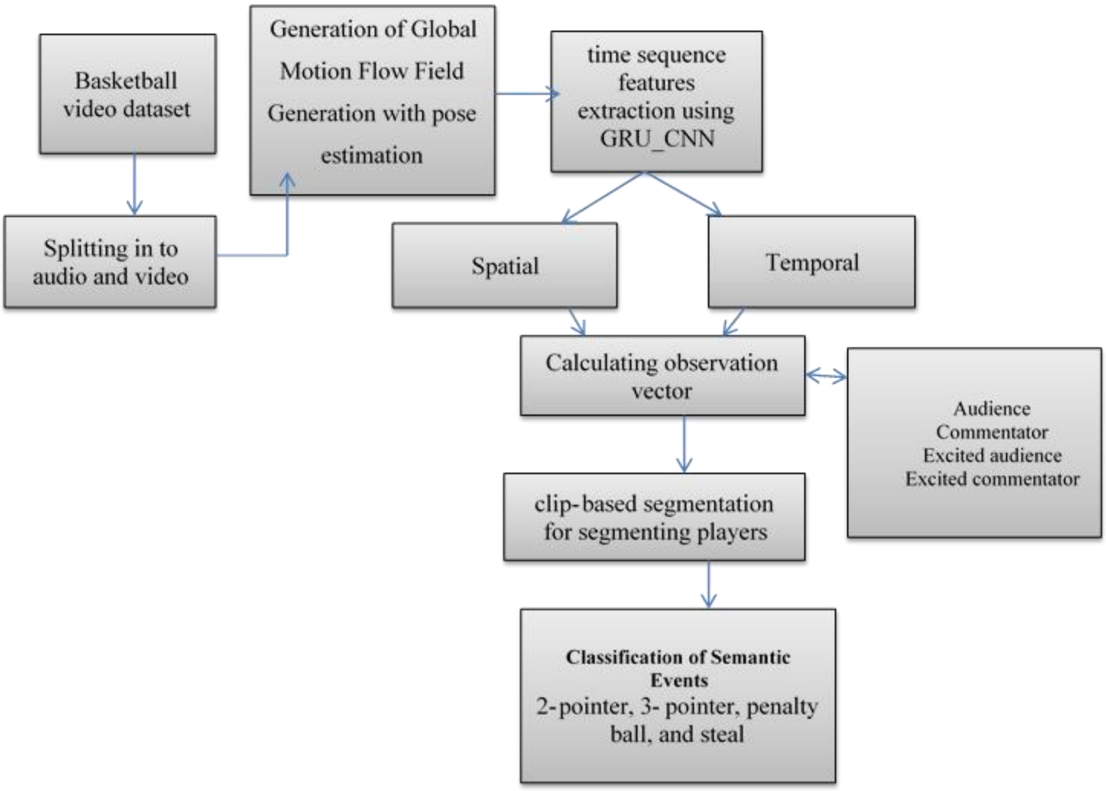
Sports have benefited greatly from the advancement of artificial intelligence, machine learning, and data mining historically. On the one hand, several techniques are suggested to forecast the results of sporting events and assist spectators in a broad analysis of sports.

[11] suggests a methodology for sophisticated healthcare management of basketball players that is video pictures cognizant of information extraction. Professional athletes' motion paths can be deduced from segmented images by utilizing a U-Net-based convolutional neural network to divide the preprocessed video images into various subgroups. Based on this, all segmentation action photographs are grouped together using the fuzzy KC-means probabilistic method into a variety of classes, where images within a class are comparable and pictures that belong to different categories are dissimilar. Throughout [12] instances, deep learning-based multi-target trajectory tracking was used in multi-frame video sequences of basketball. Each objective in the picture is detected using the YOLOv3 method in the detection phase, and the targeting characteristics are extracted using the lighter MobileNetV2 backend system instead of the previous YOLOv3 backhaul Darknet-53. The timelapse RGB map is used as the spatially source in [13] while the image sequences map is used as the chronological information. The input signals are normalized 1:1 for feature engineering. The effectiveness of the DNet basketball movement standardized recognition model recommended in this study is 94.12%, demonstrating that the strategy enhances the model's capacity to recognize while quickening learning. In order to identify occurrences in football films, [14] builds an event detection algorithm using a deep learning technique. A three-dimensional convolutional neural network is utilized to extract the frames including those inside the segment. This information is then fed into a bilateral perceptron at every time point and then further combined to produce the segment's occurrence forecast. Deep-ID network in [15] taking into account both regional and global features It creates a pose estimation recognition methodology based on graph neural network by fusing the enhanced neural network as well as the interactive and collaborative model, but then just uses the algorithm to estimate the poses of several humans. The author in [16] integrates the upgraded gray neural network method with a basketball motion video target tracking technique. The experimental test findings suggest that this system can efficiently distinguish basketball movements having high recognition rate. In [17] proposes an interactive radial basis function neural network (RBFNN). The amount of network number of hidden layers of hidden layer neurons are modified and adjusted depending on the RBF evaluation. In order to increase the precision of shot predictions in sportsThe goal of [18] is to create a massive data motion target detection system for sports complex motion image recognition using a deep convolutional neural network. Further particular, they employ the convolutional neural network's high discriminatory ability to remove pictures in order to perform computationally pretreatment for the identification of each human movement image in the video stream. The anatomical identification algorithm relies on LSTM is then utilized to recognize the essential points of the human body, which is critical for modeling various movementsIn [19], a strong template for categorizing sport visuals based on the setting and related peripheral vision is presented. Our strategy in this paper relies on the utilization of Inception V3 for extracting features and Neural Networks for categorization of different sports groups.

| Author/year | Method | Merits | demerits |
|---|---|---|---|
| Liang et al., (2023) | U-Net-based convolutional neural network | Obtain the location data in the enhanced image manually. | The lapse rate in visual data could not be captured. |
| Gong et al., (2023) | MobileNetV2 | The technology is capable of changing language features, and it is durable. | Seems to have a poor memory and cannot apply in an actual scenario. |
| Li et al., (2023) | DNet | It solves the over-smooth issue. | In the presence of unknown and complex conditions, it is impossible to forecast trajectory. |
| Liu et al., (2022) | 3D-CNN | Could solve the issue of a frontal view | Because basketball/payer behavior is dynamic, believing locations never accurately predict positions. |
| Guo et al., (2022) | Deep-ID network | Could benefit fully from relevant data | Monitoring in particular areas, such as strong background clutter, and increasing detection precision boosts reliability and calculation performance. |
| Wang et al., (2022) | improved gray neural network | capable of producing elevated voice patterns | It can reach adequately deal by using frame extracting techniques in regards to speed and accuracy, that could be a different strategic. |
| Li et al., (2021) | radial basis function neural network (RBFNN) | Quickly prepare the model | In the presence of a noisy environment, a high likelihood of obstruction, an uncommon observing position, and/or variety in movements, the suggested product's effectiveness remains inconsistent. |
| Liu et al., (2021) | LSTM | Extrapolate additional diverse characteristics | Precision can be enhanced by detecting localised changes. |
| Joshi, et al., (2020) | Inception V3 | Ability to adapt to many circumstances | When comparison to the genuine dataset, the testing set has lower identification and accuracy rate. |

## III. SYSTEM MODEL

The better prediction framework for basketball game recordings described here is founded on three - dimensional ( 3d knowledge and Global and Collective Motion Patterns (GCMPs). The two main categorization is split into four overall categories of event identification and considering two categories of occurrence identification, like depicted in Figure 1. The four main categories of event categorization were predicated on event-occ video operating segments and paired with Global and Collective Motion Pattern Deep aspects of pictures in consecutive video frame evidence, including such 2-pointer, 3-pointer, penalty ball, and steal. Audio keysounds are distinctive audio sounds that contain clear indications to major moments. Some league season audio elements (e.g., screaming, passionate commentary speaking, etc.) have significant correlation with the activities of participants, officials, announcers, and the viewers, particularly in sporting events video. Generally speaking, eager commentary speaking and enthusiastic crowd noises are key factors in sports video highlights recognition. Additional keysounds could be peculiar to a certain activity. The sound wave demonstrates successive variations in values across time, wherein parameters can be anticipated based on previous readings.


Basketball video dataset
Splitting in to audio and video
Generation of Global Motion Flow Field Generation with pose estimation
time sequence features extraction using GRU_CNN
Spatial
Temporal
Calculating observation vector
Audience Commentator Excited audience Excited commentator
clip-based segmentation for segmenting players
Classification of Semantic Events
2-pointer, 3- pointer, penalty ball, and steal

### Generation of Global Motion Flow Field Generation

Given a source picture, authors calculate dense visual features (immediate velocity) to use the regions of interest or thick illumination changes for all pixels within every frame that use the known approaches. Examine the specified frame's position $j$ Its flow matrix, $C_j$ contains the position, $A_j = (a_{j\,,}b_j)$, and the velocity, $vel_j(vel_{aj}, vel_{bj})$ i.e., $C_j = (A_j, vel_j)$. It is crucial to note that these flowing trajectories really aren't specifically connected with original image, and also no temporal sequence or entity identifiers were connected to each other. If traces are accessible but unreliable, such as shattered trajectories, flow matrices can be obtained directly from these fragmentary pathways.

Assume $\{K_1, K_2, \dots K_n\}$ The movement flow pattern is defined as $K_i = (G_i, T_i)$The sinks searching procedure phases from each location $i$ were specified as $K'_{i,t} = \left(G'_{i,t}, T'_{i,t}\right)t = 1,2 \dots$ then calculated as shown in eqn (1) and /(2)

$$G'_{i,t} = G_i, G'_{i,t+1} = G'_{i,t} + T_{i,t} \qquad (1)$$

$$T'_{i,t} = \frac{\sum_{n\epsilon neighbour\,(G'_{i,t})} T_n, H_{t,n}}{\sum_{n\epsilon neighbour\,(G'_{i,t})}, H_{t,n}} \qquad (2)$$

According to the preceding calculations, a point's subsequent 'location' is determined solely by its origin and momentum in the previous state. While the new 'velocity,' $G'_{i,t}$, is affected not just by the prior velocity, as well as the detected speeds of its contemporaries. Throughout this study, designers use kernel-based prediction, which is comparable to the active contour methodology, to integrate the local impact, while employing the continuity formula (3):

$$H_{t,n} = \exp(-\left\|\frac{G'_{i,t-1} - G_n}{h_{t-1}}\right\| \qquad (3)$$

wherein $h_{t-1}$ denotes frequency band. It should be noted that in active contour monitoring, the occurrence of images in a local region around the item is employed to estimate the object's position in the following frame.

### Extraction of time sequence features using GRU_CNN

The GRU CNN model extracts the spatial features of GMFFGs if the motion detection related to the image sequence of the movie with a duration of $T + 1$ is represented by the computation among neighboring video sequence. Throughout this system, where $\sigma$ and $tanh$ are the prominent ones, $c\langle t - 1\rangle$ is the present unit's intake, that is also the product of the previous module, and , $c\langle t\rangle$ is the present unit's result, that connects to the feedback of the subsequent component.$x\langle t\rangle$ are the classification model inputs, $y'(t)$ is the nonlinear activation output, $\epsilon_r$ and $\epsilon_u$ symbolize the resetting and refresh gates, correspondingly, and the candidates activating $cand'^{(t)}$ is computed similarly to that of the classic operational conditions. The update gate preserves past information to the present state; the value of $\epsilon_u$ ranges from 0 to 1, the closer $\epsilon_u$ is to zero, the more previous information it maintains; the reset gate is employed to establish yet if the present state and furtherance are to be merged. The quantity of $\epsilon_r$ varies between -1 to 1, with the lower number of $\epsilon_r$ ignoring greater previous information as shown in eqns (4)-(7).

$$\epsilon_u = \sigma(wei_u[c\langle t - 1\rangle, x\langle t\rangle] + bia_u) \qquad (4)$$

$$\epsilon_r = \sigma(wei_r[c\langle t - 1\rangle, x\langle t\rangle] + bia_r) \qquad (5)$$

$$c'\langle t\rangle = \tanh(wei_c[\,\epsilon_r * c\langle t - 1\rangle, x\langle t\rangle] + bia_c) \qquad (6)$$

$$c\langle t\rangle = (1 - \epsilon_u) * c\langle t - 1\rangle + \epsilon_u * cand'\langle t\rangle \qquad (7)$$

wherein $wei_u, wei_r\ and\ , wei_c$ rerepresent the learning weight matrix of the update gate, reset gate, and candidate activation $cand'\langle t\rangle$, correspondingly, and $bia_u, bia_r and\ bia_c$ crepresent the biased matrices. The spatiotemporal matrices were designed to retain the geographical data and information collected by intelligent devices and devices in the power grid, and the geospatial distance matrix information are dependent on the position of the sensor and continuity equation. The temporal matrix is represented by eqn (8)

$$\begin{matrix} z_1(1) & z_1(2) \ldots. & z_1(n) \\ z_2(1) & z_2(2) \ldots. & z_2(n) \\ z_j(1) & z_j(2) \ldots. & z_j(n) \end{matrix} \qquad (8)$$

When $j$ is the $j^{th}$ intelligent sensor, $n$ is the $n^{th}$ temporal sequences, and $z_j(n)$is the data collected by the $j^{th}$ wearable sensor at $n$ moment. CNN was employed to analyse the spatiotemporal matrices in order to remove the load feature. The aim of the convolution layers is to obtain an abstract mathematical characteristic, and the output of the convolution layer then are combined with the feedforward network. The pooling procedure doesn't at all modify the level of the input sequence, but it can minimize the size of the structure and the amount of nodes, thereby lowering the variables in the complete neural network. The vague feature was acquired and reduced to a one-dimensional vector after repeated convolution and pooling processes, allowing it to be coupled with the fully connected layer. The fully linked layer's bias and weight variables will then be determined repeatedly.

### Segmentation of players

The objective is to split the players each ten seconds, which we are able accomplish by employing our out-of-core strategy. Towards that end, researchers offer a unique clip-based segmentation technique that grows well while preserving temporal consistency without requiring the aggregate amount to be processed with one. To begin, we divide the video into equal-sized chunks of n moments (n = 26 in our studies). To maintain temporal coherence, simply add a portion (one-third) of the preceding clip's last frame to the present one. One may constrain the answer of clip $c + 1$ to be cohesive in the overlap region with clipping $c$ by adopting a grid structure and noticing that negative weighted connections usually induce a merging. Then increase the edge values $wei(edg_{a,b})$ in the transition region following creating the 3D graphs with clipping $c + 1$.

$$seg(edg_{a,b}) = \begin{cases} \delta \quad if\ Reg_{id}(p) = Reg_{id}(q) \\ 100 \times (1 - \delta\ ) \quad otherwise \end{cases} \qquad (9)$$

using $\delta \in [0, 1], \delta = 0$ for the initial period of overlap, , $\delta = 1$ for the most recent frame of crossover, and exponential somewhere between. The operator $Reg_{id}(p)$ provides the regional id from the preceding clip's segment to every voxel p in the overlapping. Like a consequence, all sides inside the same zone possess zero value, whereas all vertices between sections get a heavy strength. This compels the clip $c + 1$ segment to coincide with the clip c segment on the first frames of the overlapping, while allowing it to deviate progressively away from it in successive frames. Like a nutshell, every footage can be divided quasi-independently whilst ensuring that the categorization across all clips is chronologically constant. The accuracy of the graphic separation method is shown here. While sectioning the pictures of the participants in the movie, the estimate indices "time" and "difference in pixels" are used. The computation of the image precision is shown in Formula (10).

$$seg_{pixel} = ||U - U'||^2 \qquad (10)$$

In (10), $U$ denotes the segmented name to be evaluated; $U'$ denotes the real data label.

### Classification of Semantic Events using audio and video

Basketball situations are divided into four categories: 2-pointer, 3-pointer, penalty ball, and steal. Researchers present a heterogeneous event detection mechanism to profit both from visual and aural input in order to pinpoint the precise scenes wherein events happen. Our understand through subject matter expertise of the game of basketball that occurrence location have significant correlation with camera shake and placement. Since the video follows the participants or basketball during the game, the global utilized provides useful data for action recognition. The majority of the actions take place in settings with little camera movement. In addition, the quantity of camera movement in the next scene suggests the type of events that may occur in the current scene. To correctly assess the amount of camera motion within a picture, researchers create a feature termed altered cumulative camera motion in space ($Mod\_cm$) as the combination of ($acc\_cm$) and the dominant camera filter ($dom\_cm$),, that is,

$$Mod\_cm = acc\_cm \times dom\_cm \qquad (11)$$

Where,

$$acc_{cm} = \begin{cases} (hor_{mot} - ver_{mot}). e^{-zo_{mot}}. T_s, if\ hor_{mot}. ver_{mot} > 0 \\ (hor_{mot} + ver_{mot}). e^{-zo_{mot}}. T_s, if\ hor_{mot}. ver_{mot} < 0 \end{cases} \qquad (12)$$

where $T_s$ denotes the time period for a single scene, $hor_{mot}$ denotes camera lateral displacement, $ver_{mot}$ denotes camera vertical motion (CVD), and $zo_{mot}$denotes camera zoom. Whenever inner surface sequences share the very same advanced motion orientation inside a single shot, $dom\_cm$ is utilized to screen out another dispute.

$$dom_{cm} = \begin{cases} 1 \quad the\ first\ large\ camera\ motion\ scene\ in\ the\ long\ court - view\ shot \\ 1 \quad if\ (acc_{cm(prev)} * acc_{cm(curr)} < 0 \\ 0 \quad all\ out\ of\ court\ view\ scenes\ and\ others \end{cases} \qquad (13)$$

where $acc_{cm(prev)}$ represents the previously recognized large camera movement scenario. If $Mod\_cm$ is greater than $T_s$, this scenario features a lot of camera movement. One could divide situations into 2 groups based on the terms above. If a picture includes a significant amount of camera movement ($Mod\_cm > T_s$). then call this a both offensive and defensive interchange interval (ODI) scenario. Instead, researchers refer to it as a non-ODI scenario.

There are two types of ODI scenes depending on the sign of the $Mod\_cm$ value: ODI sequences with left-to-right image sequences and ODI sequences with right-to-left camera calibration. Some ODI situations may be missed in basketball footage since the video may focus on a single player as he or she is on the go. The identified ODI scene sequencing is enhanced further to identify such non captured ODI events. The improvement is based on the discovery that in the footage, left-to-right and right-to-left movements must swap. If there is a trial scene among two instances with the same camera movement, we allocate the latter of the two instances as an ODI picture; alternatively, we allocate the latter of the two episodes as a non-ODI incident.

## IV. PERFORMANCE ANALYSIS

The performance of our proposed carried lightweight deep learning architecture (LDLA) out by compared with three state-of-art methods such as U-Net-based convolutional neural network (U-Net CNN) [11] , MobileNetV2 [12] and radial basis function neural network (RBFNN) [17] in terms of parameters such as accuracy, precision, recall, F1-score. These parameters are calculated for four types of audio/video key sounds such as audience, commentator, excited audience, excited commentator.

Dataset description- FineBasketball [20] was created for fine-grained basketball activity recognition, and it includes three broad categories - dribbling, passing, and shooting - as well as 26 quite well categories including such backstage ball control, cross-over ball handling, hand-off, one-handed side having to pass, lay up picture, onehanded reverse layup, and block gunned down. There are 3,399 video segments in total, with each category including an average of 130 video segments.

**Table-1 comparison of accuracy and precision**

| Key sounds | Accuracy (%) | | | | Precision (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | U-Net CNN | MobileNetV2 | RBFNN | LDLA | U-Net CNN | MobileNetV2 | RBFNN | LDLA |
| audience | 78.5 | 66.9 | 71.23 | 99.8 | 67.8 | 75 | 89.5 | 91.2 |
| commentator | 77 | 67 | 73 | 98.7 | 66 | 75.8 | 88 | 92 |
| excited audience | 79.4 | 68.4 | 72.4 | 98.6 | 67 | 78 | 89 | 93.5 |
| excited commentator | 79 | 67 | 73.4 | 98.5 | 65 | 77 | 89.3 | 94 |

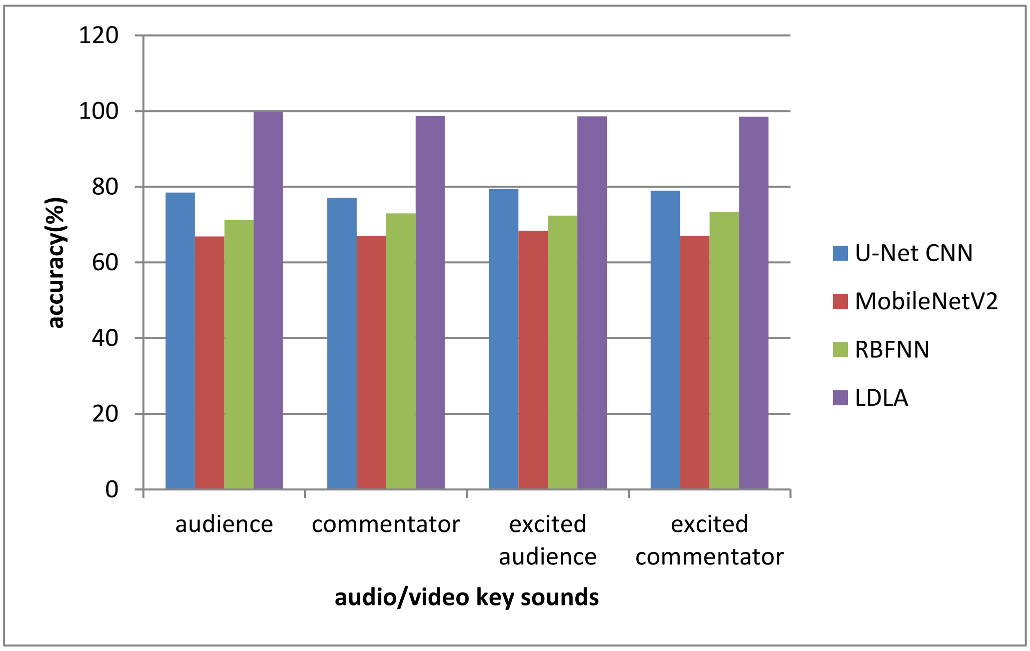


**Figure-2 comparison of accuracy**

Figure 2 depicts the accuracy comparison for the existing U-Net CNN, MobileNetv2, RBFNN, with the proposed LDLA. X axis and Y axis shows that audio/video key sounds and the values obtained in percentage respectively. When compared, existing U-Net CNN, MobileNetv2, RBFNN methods achieve 79%,69.5% and 73.2% of accuracy respectively while the proposed LDLA method achieves 99.6% of accuracy which is 20.3% better than U-Net CNN,30.1% better than MobileNetv2, and 26.2% better than RBFNN method.

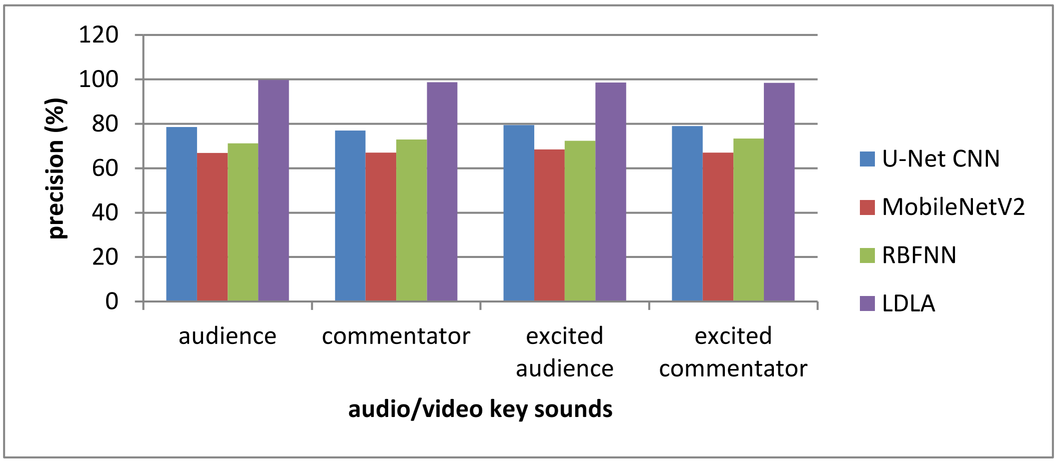


**Figure-3 comparison of precision**

Figure 3 depicts the precision comparison for the existing U-Net CNN, MobileNetv2, RBFNN, with the proposed LDLA. X axis and Y axis shows that audio/video key sounds and the values obtained in percentage respectively. When compared, existing U-Net CNN, MobileNetv2, RBFNN methods achieve 68%,78.9% and 89.4% of precision respectively while the proposed LDLA method achieves 93.2% of precision which is 25% better than U-Net CNN,15.7% better than MobileNetv2, and 4.2% better than RBFNN method.

**Table-2 comparison of recall and f1-score**

| Key sounds | recall (%) | | | | F1-score (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | U-Net CNN | MobileNetV2 | RBFNN | LDLA | U-Net CNN | MobileNetV2 | RBFNN | LDLA |
| audience | 87.6 | 71.3 | 78.9 | 91.2 | 78.9 | 67.7 | 53.4 | 89.8 |
| commentator | 89 | 72 | 78 | 92.3 | 77 | 66.8 | 52 | 89 |
| excited audience | 87.5 | 74 | 78.5 | 93 | 79.6 | 67 | 52.1 | 89.5 |
| excited commentator | 88 | 74.5 | 79 | 93.4 | 79 | 65 | 52 | 89.4 |

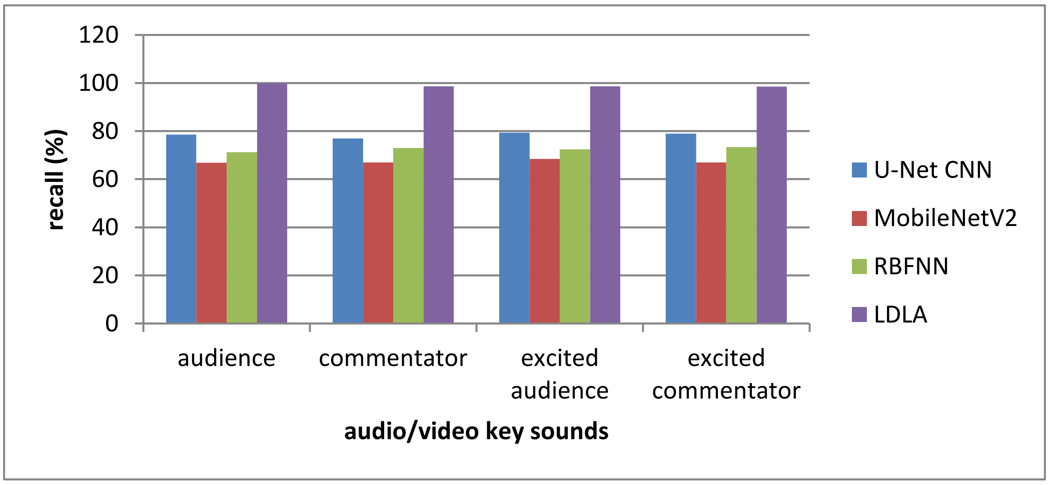


**Figure-4 comparison of recall**

Figure 4 depicts the recall comparison for the existing U-Net CNN, MobileNetv2, RBFNN, with the proposed LDLA. X axis and Y axis shows that audio/video key sounds  and the values obtained in percentage respectively. When compared, existing U-Net CNN, MobileNetv2, RBFNN methods achieve 87.4%,74.5% and 78.9% of recall respectively while the proposed LDLA method achieves 93.5% of recall which is 6.1% better than U-Net CNN,19% better than MobileNetv2, and 15.4% better than RBFNN method.

Table 3 shows the results of scene classifications for the level-1 and level-2 scenes over a total of 3,399 video segments. In the experiments, half of data set were used as training set and the remainder were used as test set.

| **classes** | **Correct classification rate (%)** |
|---|---|
| 2-pointer | 98.7 |
| 3- pointer | 97.5 |
| penalty ball | 98 |
| steal | 96.8 |

**Table-4 overall comparative analysis**

| **Parameters** | **U-Net CNN** | **MobileNetV2** | **RBFNN** | **LDLA** |
|---|---|---|---|---|
| Accuracy(%) | 79 | 69.5 | 73.2 | 99.6 |
| Precision (%) | 68 | 78.9 | 89.4 | 93.2 |
| Recall(%) | 87.4 | 74.5 | 78.9 | 93.5 |
| F1-score (%) | 78.9 | 65.4 | 53.5 | 89.5 |

## V. CONCLUSION

The knowledge management of recordings relating to sporting events has been a hot topic, in addition to the wide distribution of such videos across the internet. Previously, commentators had to discern contributing factors of players based on auditory elements and language in order to assess basketball game films. The current work explores basketball game event categorization in online footage. To begin, the group and global motion features were collected in order to convey the associated event. The current study provided a strategy for classifying basketball events based on global group motion mode and domain specific synthesis. Following that, a clip-based segment was presented to achieve automated player area segmentation via player recognition.

## REFERENCE

[1] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, A. Hilton. Computer vision for sports: Current applications and research topics. Computer Vision and Image Understanding, vol.159, pp.3–18, 2017.

[2] You, B.; Qi, H.; Ding, L.; Li, S.; Huang, L.; Tian, L.; Gao, H. Fast neural network control of a pseudo-driven wheel on deformable terrain. Mech. Syst. Signal Process. 2021, 152, 107478.

[3] Ding, Y.; Qu, Y.; Sun, J.; Du, D.; Jiang, Y.; Zhang, H. Long-Distance Multi-Vehicle Detection at Night Based on Gm-APD Lidar. Remote Sens. 2022, 14, 3553

[4] Wang, Y.; Shen, X.J.; Chen, H.P.; Sun, J.X. Action Recognition in Videos with Spatio-Temporal Fusion 3D Convolutional Neural Networks. Pattern Recognit. Image Anal. 2021, 31, 580–587

[5] Iván Alén Fernández, Fan Chen, Fabien Lavigney, Xavier Desurmontz, and Christophe De Vleeschouwer. 2010. Browsing sport content through an interactive H.264 streaming session. In 2nd International Conference on Advances in Multimedia, MMEDIA 2010. 155–161

[6] Fan Chen, Damien Delannay, and Christophe De Vleeschouwer. 2011. An autonomous framework to produce and distribute personalized team-sport video summaries: A basketball case study. IEEE Transactions on Multimedia 13, 6 (2011), 1381–1394

[7] K. C.Amit Kumar and Christophe De Vleeschouwer. 2013. Discriminative label propagation for multi-object tracking with sporadic appearance features. In Proceedings of the IEEE International Conference on Computer Vision. 2000–2007.

[8] Rudra P.K. Poudel, Ujwal Bonde, Stephan Liwicki, and Christopher Zach. 2019. ContextNet: Exploring context and detail for semantic segmentation in real-time. In British Machine Vision Conference 2018, BMVC 2018

[9] Song, Z.; Zhao, X.; Hui, Y.; Jiang, H. Fusing Attention Network based on Dilated Convolution for Super Resolution. *IEEE Trans. Cogn. Dev. Syst.* **2022**

[10] Zhao, W.; Wang, S.; Wang, X.; Zhao, Y.; Li, T.; Lin, J.; Wei, J. CZ-Base: A Database for Hand Gesture Recognition in Chinese Zither Intelligence Education. In Proceedings of the International Forum on Digital TV and Wireless Multimedia Communications, Shanghai, China, 2 December 2020; Springer: Singapore, 2020; pp. 282–292

[11] Liang, X. (2023). A video images-aware knowledge extraction method for intelligent healthcare management of basketball players. *Mathematical Biosciences and Engineering*, *20*(2), 1919-1937.

[12] Gong, Y., & Srivastava, G. (2023). Multi-target trajectory tracking in multi-frame video images of basketball sports based on deep learning. *EAI Endorsed Transactions on Scalable Information Systems*, *10*(2), e9-e9.

[13] Li, B., & Tian, M. (2023). Volleyball Movement Standardization Recognition Model Based on Convolutional Neural Network. *Computational Intelligence and Neuroscience*, *2023*.

[14] Liu, N., Liu, L., & Sun, Z. (2022). Football game video analysis method with deep learning. *Computational Intelligence and Neuroscience*, *2022*.

[15] Guo, X. (2022). Research on Multiplayer Posture Estimation Technology of Sports Competition Video Based on Graph Neural Network Algorithm. *Computational Intelligence and Neuroscience*, *2022*.

[16] Wang, T., & Shi, C. (2022). Basketball motion video target tracking algorithm based on improved gray neural network. *Neural Computing and Applications*, 1-16.

[17] Li, H., & Zhang, M. (2021). Artificial intelligence and neural network-based shooting accuracy prediction analysis in basketball. *Mobile Information Systems*, *2021*, 1-11.

[18] Liu, L. (2021). Objects detection toward complicated high remote basketball sports by leveraging deep CNN architecture. *Future Generation Computer Systems*, *119*, 31-36.

[19] Joshi, K., Tripathi, V., Bose, C., & Bhardwaj, C. (2020). Robust sports image classification using InceptionV3 and neural networks. *Procedia Computer Science*, *167*, 2374-2381.

[20] X. Gu, X. Xue, and F. Wang, “Fine-grained action recognition on a novel basketball dataset,” in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 2563–2567