Performance Analysis for Diving Sport using YoLoV8, OpenPose and Fuzzy Logic

Premanand Ghadekar
Dept. of Information Technology
VIT, Pune
premanand.ghadekar@vit.edu

Harshita Bhagat
Dept. of Information Technology
VIT, Pune
harshita.bhagat21@vit.edu

Parth More
Dept. of Information Technology
VIT, Pune
parth.more21@vit.edu

Vaishali More
Dept. of Information Technology
VIT, Pune
vaishali.more21@vit.edu

Sarthak Khare
Dept. of Information Technology
VIT, Pune
khare.sarthak211@vit.edu

Chinmay Saraf
Dept. of Information Technology
VIT, Pune
chinmay.saraf211@vit.edu

Abstract*— Diving has been an integral part of the Olympic Games since 1904.The research paper presents an innovative method for evaluating professional diving performance, a task traditionally performed by skilled judges. Through various deep learning techniques and frameworks, we propose an approach for the assessment process, which could eventually eliminate the constant need for manual judgment for athletes. Through a diverse dataset spanning various diving heights and styles, including forward and backward dives, and accounting for gender differences, our approach ensures precise performance classification. Leveraging YoloV8 for dive categorization and Open Pose for accurate pose detection and angle computation, our model enables continuous monitoring of intricate body movements. The integration of Fuzzy Logic further enhances the analysis, encompassing crucial elements such as form, execution, entry technique, and degree of difficulty. This advanced model not only facilitates precise performance evaluation but also allows for meaningful comparisons with elite divers and established benchmarks. Additionally, our meticulous creation of a tailored dataset, achieved through rigorous video segmentation, enriches the range of scenarios available for both training and evaluation purposes. In essence, this research marks a paradigm shift in the domain of professional diving performance assessment, equipping coaches and divers alike with powerful tools to refine their skills and elevate the standards of this esteemed sport.

Keywords*— Dive Analysis, Pose Detection, YoLoV8, OpenPose, Fuzzy Membership.

I. Introduction

The world of professional diving embodies a mesmerizing fusion of grace, athleticism, and sheer daring. Divers, like avian acrobats, take flight into the aquatic realm, defying gravity and executing awe-inspiring maneuvers. The evaluation of their performances, historically a realm of subjective human judgment, now stands on the cusp of a transformative era. This evolution is driven by the remarkable capabilities of deep learning methods ushering objectivity, precision, and an unparalleled understanding of professional diving excellence.

In the age of technological advancement, the harmonious confluence of multifaceted datasets, cutting-edge deep learning models, and novel analytical methodologies promises to redefine the professional diving landscape. The research embarks on an ambitious journey to revolutionize the assessment of professional diving performance, arming coaches and divers with the means not only to hone their skills but also to elevate the very standards of this venerable sport.

At the heart of this profound transformation lies the groundbreaking utilization of YoloV8 for dive categorization and OpenPose for meticulous angle tracking. These deep learning methodologies pave the way for continuous and comprehensive monitoring of body movements, transcending conventional limits in performance evaluation. The dataset, meticulously crafted, encapsulates a rich array of diving heights, spanning from 3 meters to 10 meters, diverse dive styles that encompass both forward and backward dives, and gender-specific differentiations. This extensive variety ensures the robustness and adaptability of the model, enabling it to capture the subtle intricacies of professional diving performances across a swarm of scenarios.

Further enhancing the analytical framework is the integration of Fuzzy Logic, offering a holistic approach that evaluates key performance aspects including form, execution, entry technique, and degree of difficulty. Moreover, it introduced an advanced deep learning model, skillfully interweaving OpenPose. This amalgamation orchestrates a symphony of precision, facilitating not only performance evaluation but also facilitating intricate comparisons with elite divers and established performance standards.

This paper embarks on an illuminating voyage through the annals of professional diving performance assessment, tracing its evolution from the antiquated realm of human judgment to the tantalizing possibilities of an data-driven, technologically enhanced future. By harnessing the formidable power of deep learning, the aspiration is to bridge the gap between subjectivity and objectivity, granting audiences and athletes a nuanced, data-fueled vantage point from which to appreciate professional diving's breathtaking displays. In this pursuit, the aim is not merely contribute to the sport's refinement, but also to empower athletes and coaches with the indispensable tools needed to propel themselves to new heights of excellence. This marks the beginning of a new era in how professional diving is assessed. The research shows the way to a future where diving excellence reaches new heights.

II. Literature Survey

The research by F. Vidal-Codina et al. introduces a deterministic decision tree-based algorithm is employed for the automated extraction of football events from tracking data, overcoming the limitations of manual event data collection. The algorithm achieves an impressive +90% detection rate across various tournaments, showcasing its efficacy. However, challenges related to data availability and modeling limitations are acknowledged, with proposed avenues for improvement. The study emphasizes the potential of auto-generated event data for advanced football analytics, highlighting its value in enhancing data granularity[1]. Researchers T. Murakami and T. Nakamura have addressed the challenge of estimating athlete 3D pose from monocular TV sports video with limited training data. The proposed approach utilizes a pre-trained deep neural network to estimate 3D pose from 2D joint locations, but it introduces a correction procedure to account for distortions caused by the bird's-eye view perspective common in TV sports footage. Experimental results demonstrate the effectiveness of the correction method in stabilizing 3D pose estimation, particularly for sports like tennis, table tennis, and badminton. However, the study acknowledges challenges in handling high-speed motions and suggests future work to improve accuracy and evaluate the proposed method in scenarios lacking ground-truth data [2].

The paper by E. W. Trejo and P. Yuan presents an interactive system using Kinect technology to recognize six yoga poses. The system's advantages include multimodal interaction, real-time feedback, high accuracy (above 94.78%), and expert involvement in data preparation. It offers potential benefits for yoga learners and practitioners. However, limitations such as dataset size and environmental factors should be considered, and future work may involve expanding the system to other sports disciplines and enhancing interaction options[3].Artificial immune system (AIS) was suggested by Tarek Ghoniemy et al. to obtain hyperparameter values of kernel parameters in SVM object tracking. The AIS technique performs best when improving Graph-Cut segmentation. Fast segmentation can be achieved via Graph-Cut segmentation. This paper lacks comparison with other models[4].
Muhammad Usama Islam et al. introduced a paper that utilizes Microsoft Kinect for tracking body movements and accurately assessing the correctness of yoga poses. The real-time system achieves a good 97% accuracy across all body part angles. This research has the potential for expanding applications, including recognizing a wider range of yoga poses and accommodating greater pose diversity[5].
Grandel Dsouza et al. created an AI model serving as a personal trainer for bodybuilders to assist them in reaching their fitness objectives. They employed CNN and other deep learning models in the training process. In the future, there is potential to apply this model to various medical domains. Furthermore, this foundational project could be expanded to enhance the precision of body part angles, using new techniques like OpenPose and YoLoV8[6].
Ming Xue et al. employed various deep learning algorithms to identify football shooting actions, a task of significant practical importance. In this study, the authors emphasized the significance of a feedback mechanism. The model was trained to identify observation angles that are optimal for action recognition. The findings, as reported by the authors [7], demonstrate that the network model surpasses other

models in terms of accuracy. Zihuan Shu et al.'s 2020 research and implementation of a human posture recognition algorithm using OpenPose put forth a novel approach. This approach attains an impressive recognition accuracy rate of 91.5% and operates at a speed of 9.7 frames per second. It accomplishes this by extracting vital points and constructing the skeletal representation of human posture from every individual image frame. It works well at recognizing both single and several people in a scenario and is appropriate for complex circumstances. A wide range of domains, including augmented reality, video surveillance, human-computer interaction and gait analysis, offer potential applications for the method[8].

A team from the Suzhou Municipal Hospital lead by Jing Chen reported on using the CNN-LSTM model to identify activities in video recordings of a traditional Chinese workout. The motion recognition model put forward in this work is built on long short-term memory (LSTM) and the VGG16 architecture for learning complicated features. In classifying activities, the combined CNN-LSTM model performs better than the LSTM model, reaching 96.43% accuracy as opposed to 66.07%. The suggested approach is useful for identifying the activities of Baduanjin exercises and can help with skill development. The model was evaluated using video data and has a broad range of applications in areas like rehabilitation, health monitoring, and evaluating athletic performance[9]. The study by M.R. Yeadon investigates twisting techniques employed by competitive divers during the 1991 World Student Games. Video data and computer simulations were used to predict body orientation during dives with reverse 1-somersault and 2-twists. The research found that divers primarily used asymmetrical movements of the arms and hips to generate twist, with some contribution from contact techniques. Mean deviations between simulation and video data were small, indicating the effectiveness of the study's approach in analyzing diving techniques[10].

The study by Sian Barris et al. investigates whether the kinematics (movement patterns) of elite springboard divers differ between baulked (aborted) take-offs and completed dives. They focused their analysis on the two-dimensional kinematic attributes of the take-off phases in typical training sessions.While individual variations in movement patterns were noted during the approach phase, the comprehensive analysis showed that divers did not display significantly distinct movement patterns between completed and baulked dives. This suggests that assessing the adaptability of motor behavior may be crucial for evaluating a successful dive performance, highlighting the importance of functional variability in sports[11].

This paper, authored by Gyeong-June Hahm and Keeseong Cho, presents an approach for event-based sport video segmentation. It addresses the need to summarize extensive sports video content available through CATV and OTT media services. The approach harnesses webcast textual data and extracts diverse video information from the content. Employing multimodal analysis, it identifies appropriate time intervals for each event within the video, ultimately achieving successful video segmentation, yielding promising results [12].

This paper by Stefano Frassinelli et al. presents an innovative video analysis procedure for evaluating diving performance in response to the growing trend of sport performance analysis across all athlete levels. The approach underscores flexibility and affordability, effectively tackling and solving various video processing challenges[13].

This paper by Ajay Shrestha et al. reviews various optimization methods to enhance training accuracy and reduce time in deep learning. It explores the math behind recent deep network training algorithms, highlighting current shortcomings and improvements. The survey covers a range of deep architectures, including convolution networks, residual networks, recurrent networks, reinforcement learning, and variational autoencoders[14].

This paper by Xiang Zhu et al. explores the growing significance of machine-learning techniques, particularly deep learning, in the realm of data-intensive science. It comprehensively examines the hurdles encountered when applying deep learning to remote-sensing data analysis, summarizes recent advancements, and furnishes readily available educational materials. Most importantly, it encourages remote-sensing specialists to harness the potential of deep learning as a universal framework for addressing significant issues such as climate change and urbanization[15].

III. Methodology:

The proposed model consists of 5 components:

A. Pose Detection using OpenPose:

In this component, OpenPose, a highest-rated pose detection system, is employed to accurately capture human body movements. OpenPose detects body, face, and hand other 33 keypoints, providing a detailed understanding of the diver's pose. Preprocessing techniques, such as noise reduction and keypoint refinement, are applied to enhance the accuracy of pose detection results. The model takes video as an input, converts it into frames and is passed for posture detection.

The output generated is in the JSON format, which is further converted into CSV to carry out analysis of the diver's performance.

B. Angle Computation:

In the angle computation section, joint and segment angles are meticulously calculated from detected pose key points, forming a foundational component for sports analytics and biomechanical studies. The model employs sophisticated techniques to compute joint angles based on specific key points, applying necessary offsets and multipliers for precision. Notably, segment angles are calculated with attention to anatomical accuracy, enhancing the reliability of the results.

For vector $\vec{x}$ and $\vec{y}$, to find angle between them,

$$\theta = \frac{(x \times y)}{|x||y|}$$

/

Equation 1: angle calculation from two vectors

Moreover, the code offers optional filtering methods like Butterworth, Gaussian, LOESS, and median filters to refine the angle data.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2-y^2)/2\sigma^2}$$

Equation 2: Gaussian filter for angles

Visualization options include displaying figures for comparative analysis and overlaying angles on images or videos, facilitating intuitive comprehension. The output CSV files give the angles of each attribute per frame of the given diving video.

C. Dive Classification using YOLOv8:

YOLOv8, a robust and latest object detection algorithm, is integrated into the methodology for dive classification. The system is trained on two datasets, including UCF Diving Dataset and Diving48. The integration of pose data with YOLOv8 enhances classification accuracy by providing context-aware information. Divers are classified into predefined categories, namely four: men forward dives, men backward dives, women forward dives, women backward dives, enabling a detailed analysis of dive types.

D. Dive Technique Improvement:

This component focuses on leveraging the computed angles and classification results to provide actionable feedback for divers. The data frame is divided into 3 parts: stance, flip and water entry. Angles and deviations of the dive are compared with the best dive from that segment for that local action (stance, flip or water entry). This is later compared with the threshold value, if the deviation is significant, the proposed model outputs the suggestion which can be taken into account for improvement in future dives.

Algorithm:

Input: CSV file generated from the dive.
Output: Identified deviations, associated dive elements, and corresponding suggestions for improvement.
Method: Fuzzy Logic.
Step 1: Read the diver's pose data from the preprocessed CSV files.
Step 2: Load predefined threshold values for different joint angles and segments.
Step 3: Remove rows with zero values, ensuring the dataset is clean and ready for analysis.
Step 4: Divide the cleaned data into three segments: Stance, Air Flip, and Water Entry, based on the specific time intervals.
Step 5: For each joint angle and segment in the Air Flip and Water Entry segments, calculate the mean and standard deviation.
Step 6: Compare the standard deviation with the predefined threshold values. If it exceeds the threshold, proceed with the fuzzy logic analysis.
Step 7: Check the mean deviation value. If it exceeds a certain angle threshold, provide specific feedback related to that joint angle.
Step 8: Evaluate the type of dive element associated with the deviated angle (e.g., arm position, leg alignment).
Step 9: Output tailored suggestions for improvement based on the type of deviation and dive element.

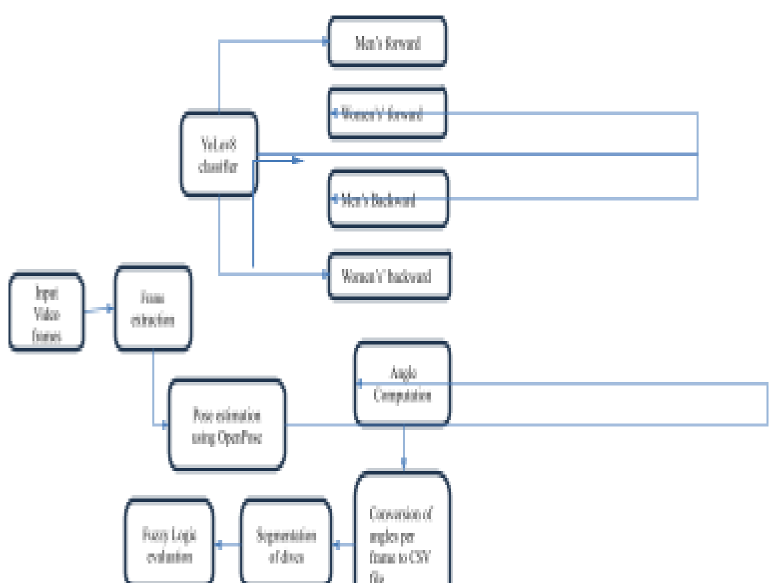

Figure 1. Flowchart for the proposed model

IV. Dataset

A. UCF Diving dataset

The UCF Sports dataset comprises a selection of sports activities frequently featured on major broadcast television networks such such as ESPN and BBC. These video clips were sourced from various stock video websites, including BBC Motion Gallery and Getty Images.

The dataset consists of a total of 150 sequences, all with a resolution of 720 x 480. It offers a diverse range of activities captured from multiple angles and scenes. By providing access to this dataset, the goal is to encourage further research in the field of action recognition in uncontrolled environments. Since its initial release, this dataset has found applications in various tasks, including action localization, action recognition, and saliency detection.

B. diving48

Diving48 is a meticulously curated video collection featuring 48 distinct dive sequences conforming to FINA standards in competitive diving. It encompasses approximately 18,000 carefully edited video clips. Contemporary action recognition algorithms face a formidable challenge in this dataset due to the inherent complexity of diving, where dives are characterized by variations in three critical stages: takeoff, flight, and entry. This complexity necessitates the modeling of long-term temporal dynamics for accurate recognition.

The video snippets in the Diving48 dataset were constructed by compiling online videos from notable diving competitions. For each dive sequence, the ground-truth labels were extracted from the information boards. The dataset is partitioned into a training set, comprising approximately 16,000 videos, and a test set, encompassing around 2,000 videos, providing a valuable resource for the development and evaluation of dive recognition models.

V. Results and Discussion

A. Classification

In this study, YOLOv8 was employed, a state-of-the-art object detection algorithm, for dive classification based on the integrated pose data obtained from OpenPose. The system was rigorously evaluated on two diverse and challenging datasets: UCF Diving Dataset and Diving48. The classification results demonstrate the effectiveness of the approach in accurately categorizing divers into predefined classes, including men forward dives, men backward dives, women forward dives, and women backward dives.

Figure 2. Mens' backward 5m dive

For the above backward dive, following is the model's output:

{0: 'Men_Forward', 1: 'Men_backward', 2: 'Women_Forward', 3: 'Women_backward'}
[0.0011691710678860545, **0.9984365105628967**, 0.00036425769212655723, 3.007608938787598e-05]
image 1/1 /Users/ucf_sports_actions/ucf_action/Diving-Side/005/4475-1_70560.jpg: 64x64 **Men_backward 1.00**, Men_Forward 0.00, Women_Forward 0.00, Women_backward 0.00, 3.0ms
Speed: 0.0ms preprocess, 3.0ms inference, 0.0ms postprocess per image at shape (1, 3, 64, 64)

As visible from the output, the model gives 99.8% (highlighted text) chances of the dive being a men's backward dive

		Diving Dataset	
model	Train loss	Val. loss	Accuracy
YOLOv8-cls	0.00228	0.00331	0.85714

TABLE 1 : YOLOv8 performance metrics

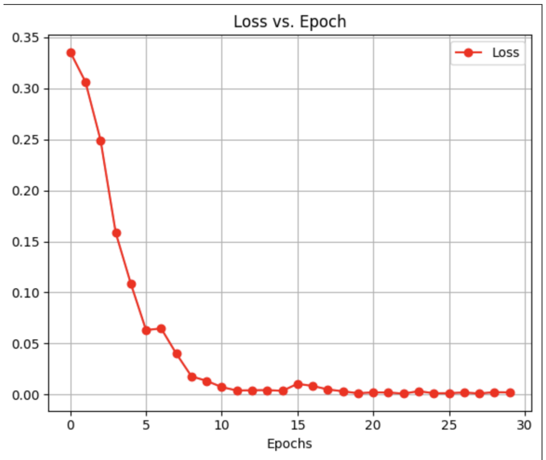

Figure 3. Loss values per epoch cycle

B. Pose Detection

The results obtained were outstanding, showcasing the system's ability to precisely identify 33 keypoints, encompassing the body, face, and hands. This detailed keypoint detection provided a granular understanding of the diver's pose throughout their performance. These key points were later used to calculate joint angles and segment angles.

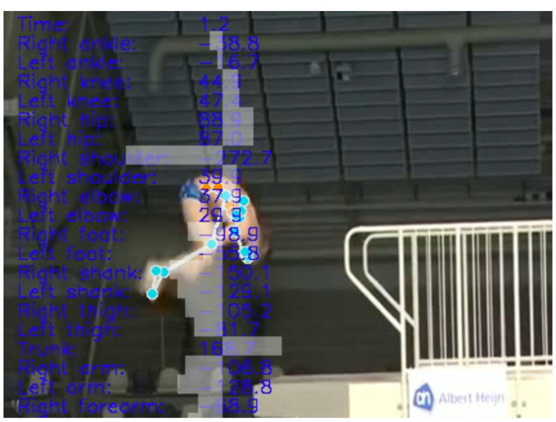

Figure 4. angle calculation

C. Dive Trainer

CSV files containing the angles of each keypoint per frame is compared. Firstly the data frame is divided into 3 parts: stance, flip, entry (in water).These 3 segments are compared with the best-in-class dive for evaluation and improvement. Deviation or wrong posture of the diver is considered if the angle detected by the model is above a threshold value, hence avoiding irregular judgment.

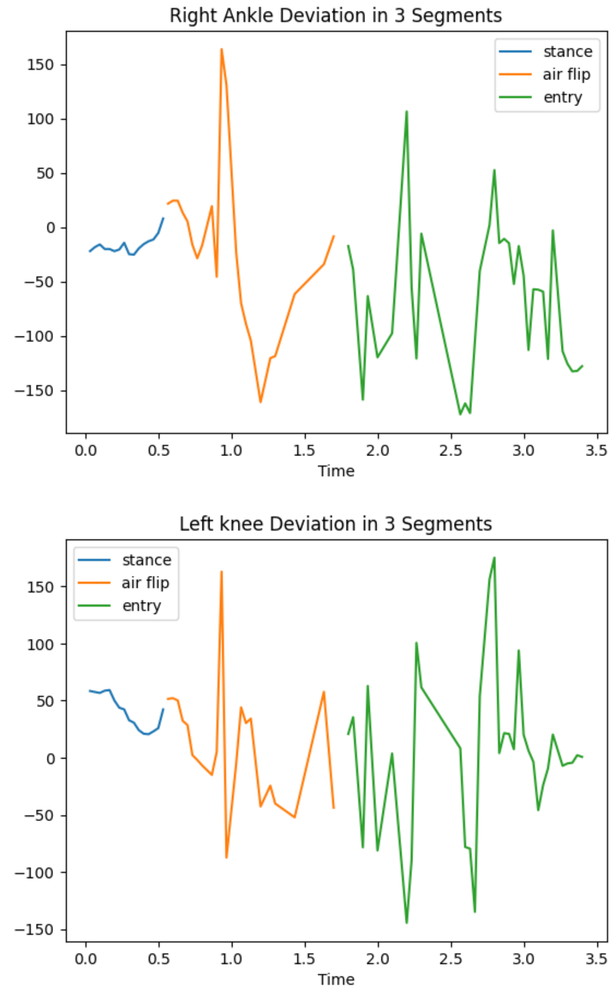

Figure 5. Deviation analysis graph for an input dive

D. Analysis

This section presents analysis of the dive performance. This performance study is divided into three sections such as take-off, air flip, and entry into the water. The analysis provides valuable insights into the divers' performance and provides meaningful conclusions for divers. For the analysis best dive of highest score is considered, input dive is compared with best dive and conclusions are drawn based on deviation in mean and standard deviation in angles of different body parts.

Implemented study uses fuzzy logic to analyze the performance of divers. Following is the membership function used in this system. The following membership function formula helps to evaluate the score of the dive.

****formula**

where,

N=number of columns, Xb(i)= best dive Mean, X(i) = input dive mean, $\sigma b(i)$ = best dive standard deviation and $\sigma(i)$ (i)=input dive standard deviation.

The above formula gives a combined deviation factor considering all features' percentage deviation in input dive from the best dive. In our approach both standard deviation and mean deviation from the best dive are considered to get better performance. This deviation factor is mapped to get a fuzzy membership function.

****membership function**

This membership function indicates that if deviation is less than or equal to deviation factor of best dive i.e. input dive is more similar to best dive, then membership of such dive is assigned as one i.e. highest score dive. If the deviation factor for a dive is greater than best dive deviation factor then that dynamic range is reduced using log function and mapped to membership values between 0 and 1.

To simplify the decision about exactly what improvement is needed and make the algorithm act as a trainer for the player, we have given analysis in 3 parts i.e. take-off, air flip, and entry into the water. Here only standard deviation will not work as you can see in **Figure(nos)** standard deviation in the take off part of both dives is same but mean deviation is different. More value of mean deviation indicates most of the time any of the body parts are deviating by significant angle, so we considered both parameters to judge performance. Mean of standard deviation and mean deviation in angles of different body parts are combined to check how much the dive is deviating from best dive, if deviation in mean and standard deviation is greater than threshold(here threshold is 50 degrees) value it means diver is significantly deviating from best dive so there is need of improvement of those body parts. Algorithm will provide a summary including deviation in angle and exactly which body parts are deviated.

Below is the analysis given for a 3m dive:

```
deviation factor:  3.43505641991592
Right ankle: deviation by: -69.71452555141177
For dives that require a tuck position, ensure a tight tuck with knees pulled close to the chest and arms wrapped around the legs. The tuck should be compact and symmetrical to facilitate faster rotations.
Point the toes and keep the legs together. The feet should be close, and ankles should be flexed slightly upward. Straight legs with pointed toes create a sleeker entry and reduce drag.
Right hip: deviation by: 34.38959960026472
Trunk: deviation by: 29.73147843852941
Right forearm: deviation by: -25.011849456529415
Keep the arms streamlined alongside the head or reaching forward. Avoid wide arm movements during entry, as they can create resistance and cause splash. Hands should be close together and fingers should be joined to minimize water resistance.
```

VI. Conclusion

In summary, this research has introduced a groundbreaking approach to the assessment of professional diving performance, harnessing the capabilities of deep learning techniques. The analysis of a diverse dataset, encompassing various dive heights, styles, and gender considerations, has significantly advanced the precision of performance classification. Employing YoloV8 for categorization, in conjunction with Open Pose for meticulous angle tracking and Fuzzy Logic for comprehensive analysis, has ushered in a more refined method for evaluating diving proficiency. Within the realm of professional diving, where a panel of judges subjectively scores each dive, the study has introduced an intricate deep learning model. This model, incorporating OpenPose, not only enables meticulous performance assessment but also streamlines the process of benchmarking against elite divers and established performance standards.

Furthermore, this study has diligently addressed the need for a bespoke dataset through systematic video segmentation, providing a diverse range of scenarios for training and evaluation purposes. This comprehensive approach ensures that coaches and divers have access to a multifaceted toolkit for honing their skills and enhancing the sport's overall standards. As deep learning emerges as a transformative force, this research marks a significant leap forward in the realm of professional diving performance assessment. The future promises a more objective, insightful, and instrumental evaluation process, benefitting all stakeholders in the sport and advancing the pursuit of diving excellence.

References