Visual Computing MAGAZiNE

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

Abstract
In this study, a deep learning model is developed for real-time and pre-recorded classification of 16 distinct fitness exercises. The model leverages the Mediapipe Pose estimation model to extract key body landmarks, which serve as training data for accurate exercise pose prediction. The extracted landmarks also enable a comprehensive analysis of the exercises, including counting exercise repetitions, providing guidance on the next move, and offering posture tips. The proposed approach demonstrates promising results in effectively classifying fitness exercises while enhancing monitoring and guidance capabilities.

1. Introduction

Regular exercise is essential for maintaining overall health and well-being. As the popularity of fitness training increases, there is a need for automated systems that can accurately classify and analyse different exercises. In this study, we develop a deep learning model to classify 16 fitness exercises. By utilizing the Mediapipe Pose estimation model, key body landmarks are extracted for training the model. Our goal is to provide real-time feedback, monitor progress, and offer personalized guidance to individuals during their fitness routines. Additionally, we perform a detailed analysis of the exercises using the extracted landmarks. This research aims to enhance monitoring and guidance in fitness training programs, promoting improved outcomes and overall well-being.

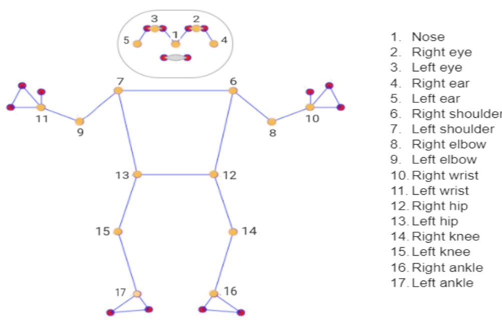

Figure 1: List of key points in our approach

2. The method

In this section, we will discuss the step-by-step process of our detailed approach. First, we will cover the extraction of 2D pose data. We will then delve into the normalization of this data and the subsequent building and training of a neural network using the extracted pose data. In the second part, we will shift our focus to the analysis of exercises based on the 2D pose data.

2.1 Exercise Recognition

Feature Extraction

In the feature extraction stage, we chose to utilize the Mediapipe Pose model [1] due to its user-friendly nature and excellent performance. The Mediapipe Pose model accurately estimates key body landmarks, such as shoulders, elbows, wrists, hips, knees, and ankles. Its ease of use and reliable performance make it an ideal choice for extracting 2D pose data from the collected dataset of videos containing 16 workout exercises.

Visual Computing MAGAZiNE

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

During the analysis of our collected videos, we observed that the key points of inner eyes and mouth did not provide relevant information for our objective of analysing and recognizing human actions. Their presence in our data did not add significant value to our data extraction process or our subsequent analysis steps. Hence, we excluded these key points from our feature extraction process, focusing solely on the essential body landmarks for exercise classification and analysis.

We extracted 2D pose data from videos of 16 exercises: 10 exercises generated by infinity.ai [2] as 3D synthetic video data and the remaining 6 exercises from YouTube videos. The distribution of the extracted 2D pose data is given by figure 2.

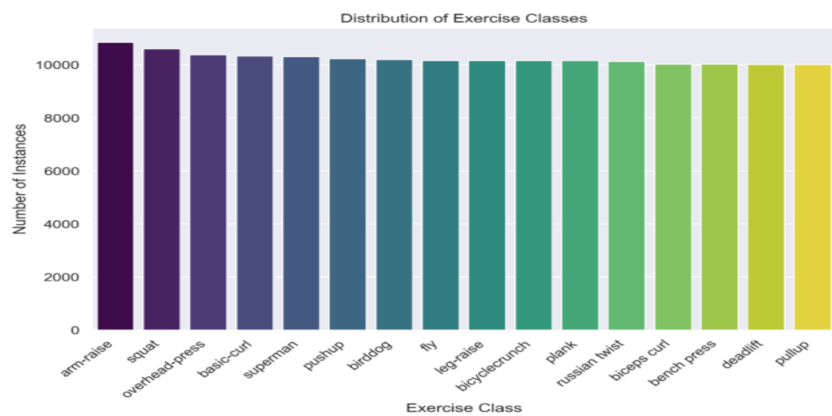

Figure 2. Distribution of extracted 2D pose data

Data Normalization

Data normalization [3] is crucial for ensuring the comparability and consistency of 2D pose data. It eliminates biases related to size and position, allowing accurate analysis and interpretation, regardless of pose or individual placement within the frame. Normalizing coordinates based on the center of gravity and body length is key to achieving this.

Model Architecture

Our neural network is designed to classify 16 different workout exercises based on 2D Mediapipe Pose data. The architecture of our model consists of several key components that enable accurate exercise recognition. The model architecture is given by figure 3.

Model Usage Method

One difficulty when using neural network models trained on Mediapipe Pose data is the rapid and fluctuating predictions. This occurs because the models are trained on static images, while the prediction task involves dynamic videos with rapidly changing poses. [4]

To address the issue, we use a queue-based smoothing technique. We store the predictions for each frame in a queue and select the most frequent prediction. This effectively smooths the prediction process and reduces rapid changes.

2.2 Exercise Analyse

Exercise analysis during execution is crucial for optimizing athlete performance by providing technical guidance and evaluating progress. The analysis includes:
-Repetition Calculation: Measures endurance progression and sets specific goals for the athlete's improvement.
-Issuing Instructions: Offers clear instructions indicating the next movement to perform during the exercise.
-Providing Relevant Tips: It provides guidance on making minor adjustments to your position, indicating if your posture is incorrect during certain exercises.
-Timer Function: Controls exercise duration for performance evaluation.

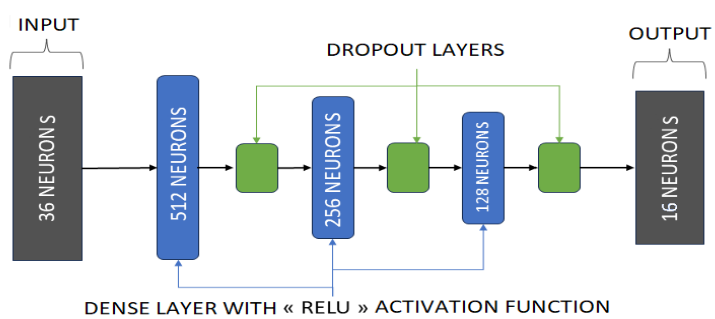

Figure 3. Model architecture

Example of Queue-Based Smoothing

For instance, in a scenario with “plank” and “push-up” pose classification, using a queue length of 5, we consider the five previous predictions. The most frequent prediction in the queue is chosen as the final prediction. Figure 4 illustrates an example. By implementing this approach with a queue length of 25, corresponding to the average frame frequency of most videos and webcams, we can mitigate the impact of rapid changes and fluctuations in predictions.

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

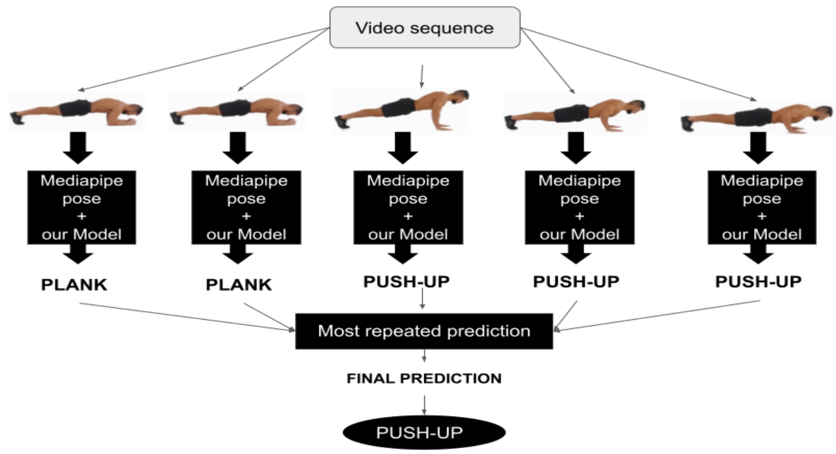

Figure 4. Queue example illustration

These aspects rely on two measures:

Calculation of Body Joint Angles

Body joint angles provide valuable information about posture, alignment, and movement dynamics, allowing evaluation of exercise techniques. The angle calculation process involves:

- Retrieving the coordinates of three key points representing crucial joints or body parts (Point A, Point B, and Point C).
- Calculating the vectors connecting these points, applying trigonometric functions to determine the angles between the vectors.
- Converting the angles from radians to degrees, and comparing them with minimum and maximum thresholds. These thresholds, experimentally established based on exercise videos produced by professionals, ensure accurate evaluation of exercise execution.

One primary application of these angles is to calculate and increase the number of repetitions of the exercise, as they serve as a measure to track progress and determine the quality of exercise performance.

Calculation of Body Joint Distances

Distances between Mediapipe Pose key points provide valuable insights into movement analysis, body alignment, and balance. The distance calculation process involves:

-Retrieving the coordinates of two key points representing the joints or body parts.
-Using these coordinates to calculate the Euclidean distance between the points.
-Comparing the calculated distance with minimum and maximum thresholds. These thresholds, experimentally established based on exercise videos produced by professionals, ensure accurate evaluation of exercise execution.

Based on the calculated angles and distances, instructions, advice, and repetition count increments are provided to the athlete. It is important to note that the specific angles, distances, thresholds, and instructions vary for each exercise and have been determined experimentally.

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

3. Experimentation

In the experimentation phase, the model underwent training and testing to assess its performance. The training process involved monitoring the model's metrics, including loss and accuracy, while the testing phase evaluated its performance on unseen data. The results were analysed using a normalized confusion matrix. Here is an overview of the experimentation process:

Training

During the training process, the model's performance was monitored using appropriate metrics. The evolution of loss, val_loss, accuracy, and val_accuracy was plotted over the training epochs. The graph below represents the metrics' progression.

The plot of figure 5 shows a consistent decrease in both loss and val_loss without fluctuations, indicating effective learning of the model. At the end of training, the loss reached 0.0296, and the val_loss was 0.0538. Simultaneously, the accuracy and val_accuracy steadily increased, reaching values of 0.9903 and 0.9847, respectively.

Testing

To evaluate the model's performance on unknown data, we carried out a test evaluation. The results are as follows:
-Test loss: 0.055
-Test accuracy: 0.984
These results indicate that the model performs well on the test set and generalizes effectively to unknown data.

References

[1] Google. MediaPipe Pose. aout 2020.GitHub, https://github.com/google/mediapipe/blob/master/docs/solutions/pose.md
[2] Brinnae Bent, « InfiniteForm: a synthetic, minimal bias dataset for fitness applications ». Medium, Nov 4, 2021, https://blog.infinity.ai/infiniteform-a-synthetic-minimal-bias-dataset-for-fitness-applications-346e69f35b81
[3] Pismenskova, Marina, et al. « Classification of a Two-Dimensional Pose Using a Human Skeleton ». MATEC Web of Conferences, vol. 132, 2017, p. 05016. https://www.matec-conferences.org/articles/matecconf/abs/2017/46/matecconf_dts2017_05016/matecconf_dts2017_05016.html
[4] Taha Anwar, « Introduction to Video Classification and Human Activity Recognition ». LearnOpenCV, March 8, 2021 https://learnopencv.com/introduction-to-video-classification-and-human-activity-recognition

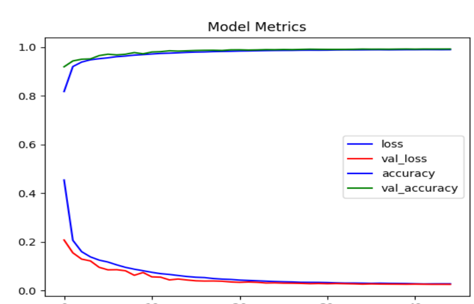

Figure 5. Evolution of the model's metrics