RECOGNITION AND REPETITION COUNTING OF PHYSICAL EXERCISES

Muhanad Tuameh, Muhammad Nasir Sabır
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, 34220 Istanbul, Türkiye
{muhanad.tuameh, nasir.sabir}@std.yildiz.edu.tr

***Özetçe* —Spor eğitim videolarında insan poz tanıma için Poz Tahmini tekniğini kullanma hedefine ulaşmak için görüntünün hangi egzersizi temsil ettiğini belirleyebilmek ve tanımlayabilmek önemlidir. Bu, sistemin belirli bir egzersize karşılık gelen farklı pozları ve hareketleri anlayabilmesi gerektiği anlamına gelir. Bu, sistemin belirli bir egzersiz hareketinin tekrar sayısını doğru bir şekilde algılamasına ve saymasına izin verdiği için süreçte önemli bir adımdır. Verilen görsel verileri bilgisayar tarafından anlaşılabilir bir sunuma dönüştürmek için MediaPipe çerçevesini kullandık. Bu çerçeve insan vücudunu 33 ayrı nokta ve koordinatlarla işaretleyerek görsel bir veriye dönüştürebilmektedir. Daha sonra makine öğrenimi ve derin öğrenme tekniklerini kullanarak pozu tahmin etmeye çalıştık. Kullanılan yöntemler Rastgele Orman ve Çok Katmanlı Perceptron Sinir Ağı'dır.**

Anahtar Kelimeler—Egzirsiz Tespiti , Makine Öğrenmesi,MediaPipe, Spor.

***Abstract*—In order to achieve the goal of using the Pose Estimation technique for human pose recognition in sports training videos, it is important to be able to identify and define which exercise the image represents. This means that the system needs to be able to understand the different poses and movements that correspond to a specific exercise. This is an essential step in the process, as it allows the system to accurately detect and count the number of repetitions of a specific exercise movement. We used the MediaPipe framework to transform the given visual data into a computer understandable representation. This framework can transform the human body into a visual data by marking it with 33 separate points with coordinates. Then we tried to predict the pose using machine learning and deep learning techniques. The methods used are Random Forest and Multilayer Perceptron Neural Network.**

Keywords—Pose Detection ,MediaPipe, Machine Learning, Sport.

I. INTRODUCTION

Human pose recognition is an important area of research in computer vision that has the potential to be applied in various fields such as sports training, and human-computer interaction. In this project, we are going to use the Pose Estimation technique to recognize and analyze the movements of athletes during training. By using machine learning and deep learning techniques, the proposed project aims to develop a solution that can accurately detect and count the number of repetitions of a specific exercise movement, providing instant feedback to athletes and coaches. This can help optimize training regimens and increase the efficiency of the overall training process. Before the age of technology, the fitness industry, like many other industries, had to make its calculations based on human skills. Even in professional work environments, it was not easy enough to follow the progress of the athletes and get high efficiency from their training.

In this project, our goal is to use the Pose Estimation technique to define the movements of the athletes in the training videos given to the program by using machine learning and deep learning techniques, and then to produce a solution that can count the number of repetitions of the defined exercise movement and display the output of it in an instant. This development provides a solution against the counting errors that the training athletes can make while tracking the number of repetitions, and it is aimed to ensure that they focus more on the movement accuracy of the exercise instead of counting the exercise.

II. RELATED WORK

Human pose recognition is a rapidly growing field of research that has attracted significant attention in recent years due to its potential applications in sports training, rehabilitation, and human-computer interaction. Several studies have been conducted in this area to develop techniques for accurately detecting and analyzing human poses.

One of the earliest approaches in this field is the use of template matching algorithms, which compare the input image to a set of predefined templates of human poses. These methods, however, are limited in their ability to handle variations in pose and body shape.

Recent studies have focused on the use of wearable devices to collect the data required for the classification, These methods have shown promising results in terms of accuracy and robustness to variations in pose and body shape. For example, the study by Ishii et al. (2021) [1] proposed an approach based on 4 wearable devices that can be mounted on different parts of the body. The approach achieved a good performance on for most of the regular exercises. Similarly, the study by Hussain et al. (2022) [2] proposed an LSTM-based approach that achieved high accuracy on several exercises.

However, despite the progress made in the field, there are still some limitations that need to be addressed. One major limitation is the need of wearable devices, which can be hard to use and expensive.

In conclusion, the literature on human pose recognition has made significant progress in recent years, with deep learning techniques showing great promise. However, there are still some limitations that need to be addressed, such as occlusion and real-time performance. Future research should focus on addressing these limitations and continuing to improve the accuracy and robustness of human pose recognition methods.

III. Data-Set

This dataset represents 10 different physical poses that can be used to distinguish 5 exercises. The exercises are Push-up, Pull-up, Sit-up, Jumping Jack and Squat. For every exercise, 2 different classes have been used to represent the terminal positions of that exercise (e.g., “up” and “down” positions for push-ups).

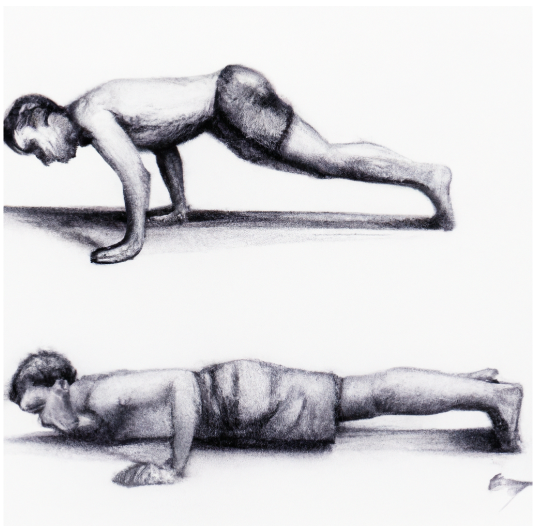

Figure 1 Terminal Positions for Push-up exercise [3]

About 500 videos of people doing the exercises have been used. The videos are from Countix [4] Dataset that contain several human activity videos. From every video, at least 2 frames are manually extracted. The extracted frames represent the terminal positions of the exercise.

For every frame, MediaPipe framework [5] is used for applying pose estimation, which detects the human skeleton of the person in the frame. The landmark model in MediaPipe Pose predicts the location of 33 3D pose landmarks (see Figure 5.2).

Instead of just using the locations of the body landmarks, we attempted to optimize the dataset by extracting 2 different kinds of features from the 33 pose landmarks:

1) Distances: The distances are calculated between specific pose landmarks using the 33 pose landmarks. Since we have the 3-dimensional location of the landmarks, we calculated the distances in 2 different ways; First, we calculated the distances between the points separately for each dimension and saved the distances in a table named “xyz-distances”. Thereafter, the 3D distances between the same points are calculated and saved in a table named “3d-distances”.

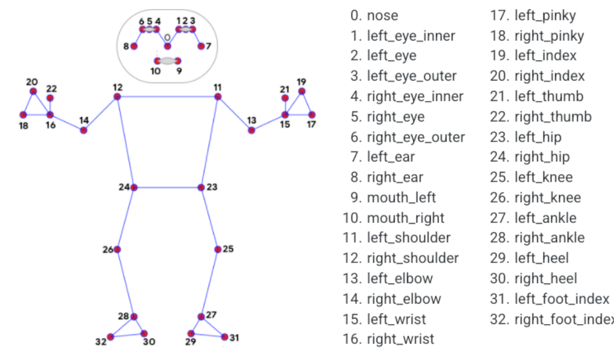

Figure 2 The detected 33 landmarks[6]

2) Angles: The angle measures of 7 different points of the human body are calculated. The angles are measured in degrees and saved in a table named “angles”.

The landmarks and the extracted features are shared publicly on Kaggle.

IV. Methodology

The purpose of the project is to build an AI model that can recognize a fitness pose and count the repetitions of that fitness exercise in real-time. The input of the model is a video of a human doing the exercise, and the output is the name of the exercise and a number of how many times the exercise have been done. The repetition counter should update the number of the repetition based on the change of the pose during the video.

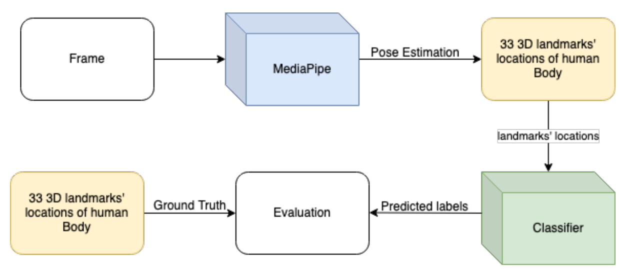

Figure 3 Block Schema of The Proposed Pipeline

In order to create the described model, different machine learning and deep learning techniques has been used and compared to each other, so we can select the best performing method in our model. To be able to train the AI model, a dataset of 1370 images of 5 fitness exercises are collected in the first step of the project. The plan is to train the model to recognize the terminal positions of every exercise, so it will be able to count the repetitions based on them e.g. if the exercise performer changed his/her position in a particular sequence (i.e. up, down, up or down, up, down) the model will consider it as a repetition of the exercise. Also, if the model learned the terminal positions of every exercise, it would also be able to recognize the exercise.

The collected dataset contains small amount of images. It is impossible for an AI model to learn the exercise poses by just this small amount of raw images. Thus, pose estimation technique is used to detect the human body skeleton. MediaPipe framework is used for pose estimation; MediaPipe Pose perform body pose tracking by inferring 33 3D landmarks. This way, the visual data is converted into numerical data that can be fed to the AI model.

On the webpage of MediaPipe K-nn algorithm is used for the purpose of pose classification. Although the results of the method were not very bad, the proposed method needs to specify the target pose to be able to count the repetition, which is not efficient from the user perspective.

Instead, we implemented 2 methods that can give us more accurate classification results and without the need of specifying a target pose.

1) Random Forest
2) Multi-Layer Perceptron Neural Network

V. Results

In this project, two methods were used for human pose recognition: Random Forest and Multi Perceptron Neural Network. Both methods were trained and tested on a dataset of sports training videos. The Random Forest method achieved an accuracy of about 82%. It is an ensemble learning method that combines multiple decision trees, which makes it robust to variations in the data. However, it was observed that the method had a slight overfitting issue, meaning that it performed well on the training data but slightly worse on the test data. The Multi Perceptron Neural Network achieved an accuracy of 84%. It was observed that this method showed slightly better performance than random forest and had less overfitting issue.

A. Random Forest

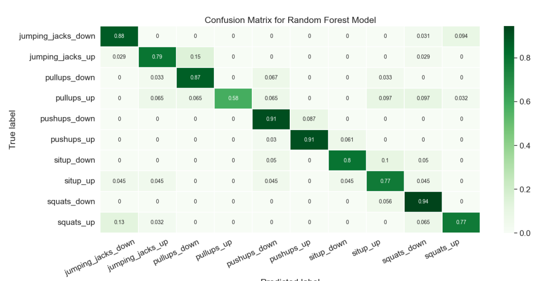

Figure 4 The confusion matrix of the Random Forest model

B. Multi Perceptron Neural Network

VI. Future Work

Our current model was limited by the small amount of data used and the results were based on only two cases, which resulted in lower accuracy rates for certain exercises. We trained the body movement by dividing it into only two states, however, this can be improved in future by using all

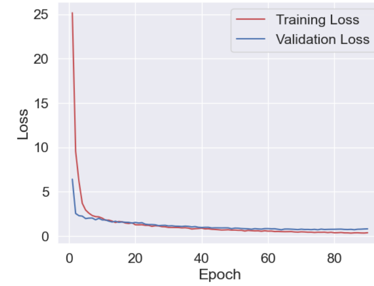

Figure 5 Loss during training

Figure 6 Accuracy during training

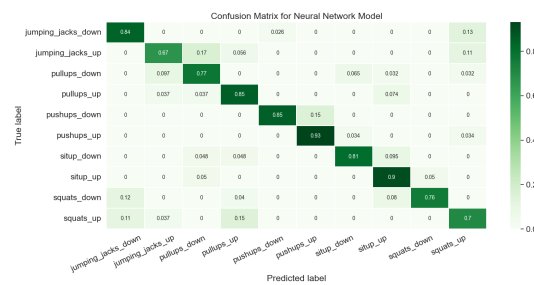

Figure 7 The confusion matrix of the Neural Network model

the sequential moves in an exercise as time series data to train a model.

In order to improve the performance, we plan to use transformer or Recurrent Neural Network (RNN) such as LSTM, on the time series data during the training stage. Instead of using the terminal state for counting the number of repetitions, it's possible to convert the changes of some parts of the human body into a signal and use the periodic nature of the move to count the number of repetitions. This approach is likely to significantly improve the results.

VII. Conclusion

In conclusion, the project aimed to use the Pose Estimation technique for human pose recognition in sports training videos in order to improve the performance and safety of athletes. Machine learning and deep learning techniques, Random Forest and Multilayer Perceptron were used to detect human poses. The results were promising, showing that the use of these techniques can usable to get a high accuracy and robustness of human pose recognition.

The main results obtained from the project were the development of a solution that could accurately detect and count the number of repetitions of a specific exercise movement, providing instant feedback to athletes and coaches.

Overall, the project successfully achieved its target, however, there are still areas for improvement. It is recommended for future work to focus on addressing these limitations and continue to improve the accuracy and robustness of human pose recognition methods. Additionally, it would be interesting to explore other applications of the developed solution, such as in rehabilitation and human-computer interaction.

References

[1] S. Ishii, A. Yokokubo, M. Luimula, and G. Lopez, "Exersense: Physical exercise recognition and counting algorithm from wearables robust to positioning," *Sensors*, vol. 21, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/1/91

[2] Z. K. Hussain A, A. L. Baig AR, Almakki R, and K. S, "Sensor-based gym physical exercise recognition: Data acquisition and experiments." *Sensors*, vol. 22, no. 7, 2022. [Online]. Available: https://doi.org/10.3390/s22072489

[3] DALL·E. (2022) 2 human bodies when one represents the up position of push-up exercise and the other represents the down position of push-up exercise, pencil art. [Online]. Available: https://labs.openai.com/s/v0XUp9wsYX2tilbroMBJcsic

[4] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, and A. Zisserman, "Counting out time: Class agnostic video repetition counting in the wild," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[5] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," 2019. [Online]. Available: https://arxiv.org/abs/1906.08172

[6] Mediapipe pose. [Online]. Available: https://google.github.io/mediapipe/solutions/pose.html