FRAMEWORK DESIGN FOR KEYPOINT EXTRACTION AND ANALYSIS BASED ON DEEP LEARNING FOR THE IMPROVEMENT OF THE TECHNIQUE USED IN PHYSICAL ACTIVITIES

Author: Hernández Bas, Ignacio.
Supervisor: Clemente Verdú, Álvaro.
Collaborating Entity: BigML, Inc

ABSTRACT

The *Deep Learning* and *Pose Detection* branches are rapidly evolving, but despite this progress, their utilization is quite complex. To facilitate their use , an application has been developed to enable the detection and analysis of the technique used during the practice of different physical activities. By the creation of a dataset and evaluating the system over this data, satisfactory results have been obtained, demonstrating the potential of pose detection in the field of sports.

Key words: Deep Learning, PoseDetection, keypoints

1. Introduction

During the last years, the branch of Deep Learning within the field of Machine Learning has risen in popularity considerably due to the continuous development of technologies and the emergence of new techniques and research. Deep Learning is defined as a set of techniques that are based on algorithms that aim to replicate the structure and functioning of the human brain with the creation of Neural Networks. Various of the most popular applications are autonomous driving, language modeling, Natural Language Processing (NLP) and image processing and classification.

The advancement in image processing techniques has led to a need for event classification, such as the movements or poses performed by an individual. The branch responsible of performing this kind of detection is referred as Pose Detection, applying techniques where computers attempt to replicate how humans assimilate information through their eyes. Pose Detection focuses on the creation and design of models that can locate the keypoints of the entity that is performing the action (e.g., a human). These keypoints refer to the body parts detected by the pose detection model, which returns the spatial coordinates of each of these points at a specific moment in time.

Once the coordinates from the model have been obtained, other systems can use these results and combine them with other types of information, such as the context or the type of activity that is being performed to obtain a deeper understanding of the action of the user.

Despite the existence of application that already incorporate pose detection techniques such as the detection of possible mistakes on during yoga practice[2] or analyzing a golfer's swing [3], these applications are often limited to simple analyses that provide limited information to the user. In addition, these applications are only focusing a narrow range of specific actions.

2. Project definition

This project defines the creation of a system that allows bringing to the user multiple functionalities and tools related to the detection and analysis of poses in the field of sports and physical exercise. Therefore, an application for the detection and analysis of physical activities will be designed, in which the user will be provided with the information related to his movements and the details to carry out the necessary corrections in case he is not executing them correctly according to the type of exercise he has performed.

This project is mainly focused on the detection of physical exercises that a user could perform either at home or in the gym and with the possibility of using different objects such as weights or bars to carry out their activity.

3. System description

The system created seeks to reduce the level of understanding necessary to be able to make use of the benefits of pose detection by creating a highly usable tool that can effectively and accurately detect events that occur during physical activity. In order to carry out this function, the system is divided into three fundamental sections, which are represented in figure 1.

- **Web application:** the web interface created in the system allows the user to interact with the system, letting the user upload videos for processing, analysis and representation of results.
- **Pose processing and detection**section of the system in charge of ingesting the video introduced by the user, carrying out the detection of the user's pose in each of the images (frames) that compose the video. This detection will be performed based on the MoveNet [4] open-source Pose Detection model.
- **Event analysis and retrieval**: Once all the images of the processed video with all the poses detected are obtained, the analysis of each of these poses is performed according to the type of physical exercise that the user has performed. Each type of exercise implies that a series of specific rules must be applied to achieve the target pose that is considered correct. This is done from the grouping of a series of general rules, which encompass the movement and position of a set of particular body parts such as legs, arms, or torso..

As a result, there is an adaptable system that allows the incorporation of new physical activities without the need to generate specific rules depending on the activity to be incorporated into the application. After the analysis, the results are summarized in unique events that provide useful information to the user where the actions that have been performed in the video are defined.

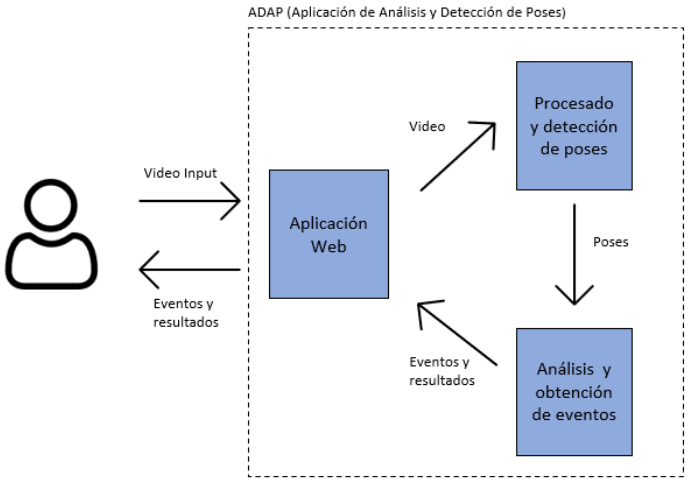

Figure 1 Principal sections of the system

4. Results

After the implementation, it is important to evaluate the behavior of the system, determining whether the predictions made by the analysis section are being carried out correctly.

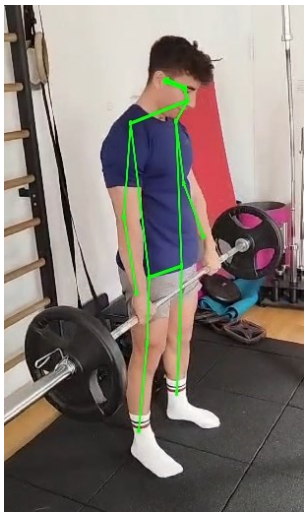

Figure 2 Objective position has been reached

Figure 3 Objective postion has not been reached

Figures 2 and 3 show examples of the images resulting from the process of analysis and detection of a video. These representations serve as indicators to the user about the performance obtained in the activity. The different parts of the body are highlighted in green or magenta, depending on whether they are within or outside the target range of the pose to be achieved based on the activity.

By creating a dataset where multiple different physical activities have been labeled, it has been possible to measure how accurate the system is in detecting the different physical activities..

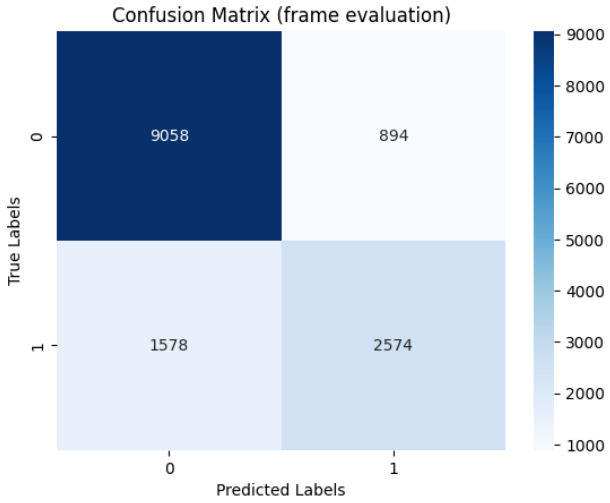

Figure 4 Confusion matrix of prediction results

Métrica	Valor
Accuracy	0.82
Precision	0.74
Recall	0.61
F1 Score	0.67

Table 1 Evaluation metrics obtained

Figure 4 shows the confusion matrix resulting from the comparison between all the labeled images and the images predicted by the system. The accuracy results show that more than 80% of the labeled frames are correctly predicted, other metrics such as precision indicate that 74% of the times an error-free pose is detected, it was also rated as correct.

On the other hand, recall exceeds 60%, which reflects that the system generally detects correctly the frames with a correct pose, but there are cases that the system does not detect. In view of these results, it can be stated that the system's performance is satisfactory, but there is room for improvement in its predictions.

5. Conclusions

Thanks to the creation of this project, it has been possible to present to the user the utilities offered by the detection of poses through sport, opening a new door to an innovative way of exercising.

The results obtained corroborate the good performance of the system, but it is of great importance to mention that the amount of data and the variability between them is a determining factor when carrying out an evaluation. The more data and situations you have, the more information you can obtain from the evaluation of the predicted data, obtaining better conclusions about the performance of the system. Therefore, it is recommended to take the established dataset as a useful starting point to determine the current accuracy of the system.

Regarding possible future work, it would be interesting to compose a specialized dataset within the framework of the project, which would present a much larger extension and concentrate numerous physical activities, performed by many people and from a greater number of different perspectives than those presented by the current dataset.

On the other hand, it is recommended to investigate Pose Detection methods that incorporate multiple cameras when recording the activity. Although MoveNet is a model that provides good results with the use of a single camera, a three-dimensional analysis of the action could improve the detection of the user's key points and therefore obtain greater accuracy in the analysis of physical activities.

6. References

[1] ABDULLAH, Malak; MADAIN, Alia; JARARWEH, Yaser. ChatGPT: Fundamentals, applications and social impacts. En *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 2022. p. 1-8.

[2] Ajay Chaudhari et al. YOG-GURU: REAL-TIME YOGA POSE CORRECTION SYSTEM USING DEEP LEARNING METHODS.

URL:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9509937.

[3] Marc Marais and Dane Brown. Golf Swing Sequencing Using Computer Vision. URL: https://link.springer.com/chapter/10.1007/978-3-031-04881-4_28

[4] Google. MoveNet: modelo de detección de pose ultrarrápido y preciso. URL https://www.tensorflow.org/hub/tutorials/movenet.