Physical exercise form correction using video processing

Sachin K Rao[1], Rohan Savalgi[2], Vineet K S[3], Nikita Pyati[4], Sunitha K[5]
Department of Information Science and Engineering
RNS Institute of Technology

[1]orcid.org/0000-0001-9142-9045
[2]savalgirohan@gmail.com
[3]vineetks2001@gmail.com
[4]nikipyati@gmail.com

***Abstract*—The importance of the way human beings conduct their health can result in numerous benefits which include advanced cardiopulmonary capacity, concentration, and prevention of weight problems, that may successfully lessen the threat of loss of life. Working out at home is an opportunity to make fitness accessible without requiring lots of equipment. However, it has its drawbacks such as a confined access of fitness know-how, which can lead to severe injuries. Therefore, an accurate fitness detection system is vital to enhance the health consciousness of human beings. Our solution uses the MediaPipe library which is trained on the InfiniteRep dataset and fine-tuned for real-time form detection and correction. By using this product, fitness enthusiasts can get accurate and real-time information about their exercise technique, which can help them improve performance and prevent injury.**

***Index Terms*—Video Processing, Physical exercise, Machine Learning, Computer Vision, MediaPipe.**

I. Introduction

This research focuses on developing a real-time system to detect and correct physical exercises which mainly include lunges, bicep curls, push-ups and leg extensions using video processing techniques. It also aims to improve the safety and effectiveness of these physical exercises by providing real-time feedback on users' form and posture. Posture correction is essential for maintaining good health and preventing various health problems. Poor posture can lead to muscle imbalances, joint pain, and decreased mobility, among other issues. One of the main benefits of proper posture is that it can help prevent back pain.

When we sit or stand with poor posture, it places excess stress on our back muscles and spinal discs, leading to discomfort and pain. Correcting posture can relieve this stress, reducing the risk of back pain. Correct posture can also improve breathing and digestion. When we slouch or hunch forward, it compresses the chest and abdomen, reducing lung capacity and affecting digestion. Good posture allows the lungs and diaphragm to function properly, promoting healthy breathing and digestion.

The dataset used for training is *InfiniteRep*, which is a synthetic, open-source dataset for fitness and physical therapy (PT) applications. It includes 1000+ videos of diverse avatars performing multiple repetitions of common exercises. It includes significant variation in the environment, lighting conditions, avatar demographics, and movement trajectories. The dataset features 100 videos per exercise, spanning 5 to 10 repetitions each (1,000 videos total) . It also includes 7 unique indoor surroundings by providing realistic environmental conditions along with their own corresponding labels. The main features include working under diverse lighting conditions, varied body shape, skin tones, and clothing. The dataset can also be used for rich annotations for 2D and 3D supervision.

To carry out this work, we used two main libraries, Mediapipe and OpenCV. Mediapipe is an open-source framework for building multimedia pipelines developed by Google, while OpenCV is a library of computer vision functions used for image and video processing, initially developed by Intel. The powerful combination of these two libraries provides an efficient and useful toolkit for detecting and analyzing human body movements and positions in real-time. Video footage of users performing exercises is captured using OpenCV and their movements are analyzed using Mediapipe. Figure 1 shows the 33 body landmarks detected using Mediapipe for pose estimation.

The model then compares these movements to a database of correct exercise forms and provides feedback on areas where the user's form needs improvement. The project has the potential to be used in a variety of settings, including gyms, fitness classes, and personal training sessions. It can also be customized to provide personalized feedback based

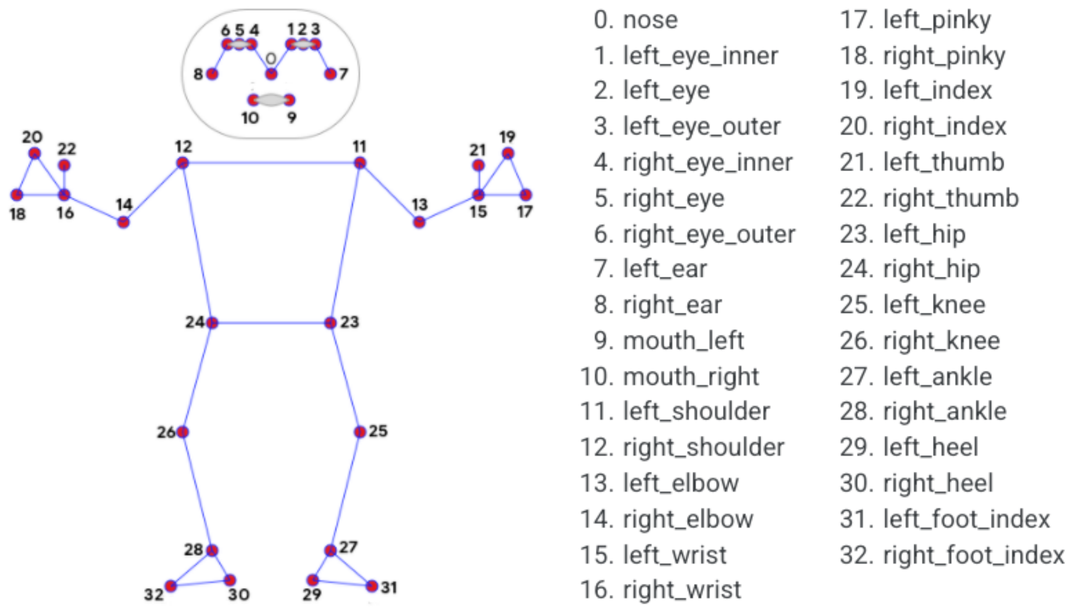

Fig. 1. Body landmarks detected using MediaPipe

on users' experience levels and fitness goals.

This system is trained on the InfiniteRep dataset before being used to provide feedback to the user on areas where their form needs improvement, such as adjusting their posture or modifying their movements. The feedback is presented in a clear and intuitive way, using visual cues and text prompts.

II. Related work

C. Militaru et al. [1] created a database of 2400 images to train a CNN to classify images to monitor and correct the posture during physical exercises. Since there were many apps in the market showing all the postures for each exercise, they often require expensive hardware.

Y. Agarwal and A. Sharma [2] obtained a similar analysis that stated that there is an increasing popularity for a way to enhance bodily and mental well-being. To answer this growth, they created a dataset containing a minimum of 5500 snap shots of ten unique yoga poses. Angles of the joints in the human body are extracted using the tf-pose skeleton and used them as a feature to implement various machine learning models.

A. Badiola and A. Mendez [3], wrote a paper discussing Human Pose Estimation (HPE) and the performance of Deep Learning and its applications in sports and physical exercises. The aim of the authors was to create an analysis regarding application of HPE in SPE, available data, methods, performance, opportunities, and challenges. One reviewer used different inclusions and exclusions criteria, as well as quality metrics. Following the authors' methods, the use of universal systems such as foundation such as OpenPose combined with other methods and tailored to a specific use case to be found.

A. Hannan, M. Shafiq, F. Hussain and I. Pires [4] published a paper which talks about how fitness and sports have attracted significant attention in the field of wearable and compelling computing. Since working out has a lot of benefits attached to it, the authors discovered problems such as sudden movements and incorrect posture during training can lead to temporary or permanent disability. In order to overcome these serious problems, the authors proposed fitness set is equipped with a gyroscope and EMG sensory modules for performing T-bar and biceps curl exercise. It provided alerts for unhealthy, incorrect posture movements through an Android app and guided users to the best possible posture based on sensor readings, wherein they used a KNN classification model to predict and guide the user to perform a specific exercise using a virtual gym trainer through a text-to-speech engine.

Although there has been some prior research done in the area of physical exercise form correction, most systems either use wearable devices which impact the user's experience and potentially their performance, or they only perform form correction for static images, which cannot provide real-time insights to the users. The aim of this work is to overcome these limitations through a system that can process real-time video and instantaneously provide form correction insights to the user.

III. Proposed System

The proposed system for this work uses a webcam to capture the user's movements, and a trained MediaPipe model to estimate their body pose in real-time. The system can then provide feedback to the user on how to correct their form to achieve a better workout. Figure 2 shows the workflow of the project starting from the acquisition of the InfiniteRep dataset, which is a large dataset of weightlifting exercises captured from multiple camera angles and stored in an MPED-V AVC (.mp4) format. The model is first trained on the data and fine-tuned based on its performance in order to be able to accurately apply it to a real-time webcam video feed.

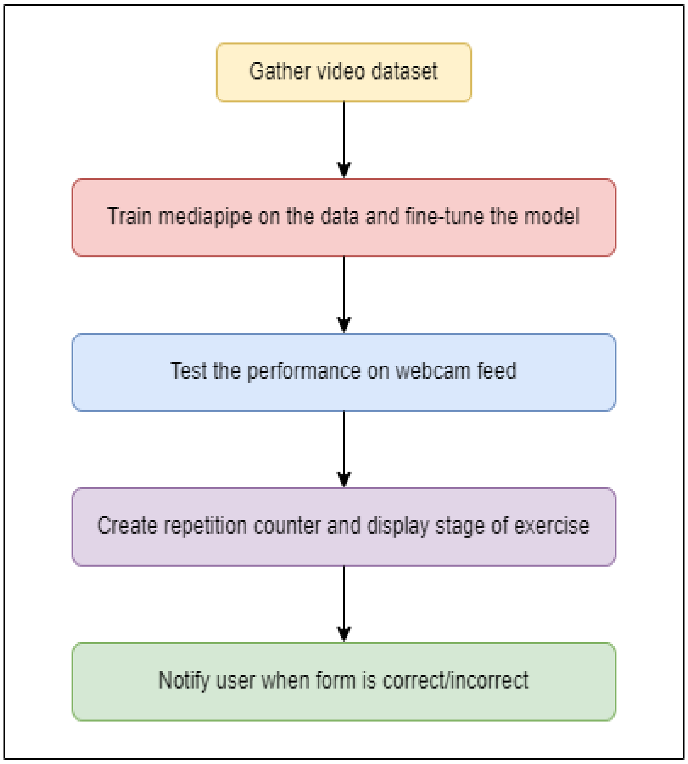

Fig. 2. System architecture

The estimated joints and body parts are then checked for correctness by the use of functions which calculate the angles between them based on the specific exercise being performed. This is done by calculating the element-wise arc tangent of two co-ordinates, x and y in the form of *(x/y)*, by choosing the quadrant correctly. The quadrant, or in other words, the branch is chosen such that the arc tangent of *(x,y)* is the signed angle in radians between the ray ending at the origin and passing through the point *(1,0)*, and the ray ending at the origin and passing through the point *(y, x)*. This value in radians is then converted to degrees and altered such that any angle greater than 180° is reset to 0°. This is done to calculate the variation of the posture from the accurate form, instead of just obtaining the numeric value without being able to use it to derive insights.

For the purpose of this work, we have considered four different exercises, which are push-ups, leg extensions, lunges and bicep curls. If the user's form is incorrect, the system provides real-time feedback on where the user is making a mistake, and what can be done to correct it. This feedback is in the form of visual cues such as highlighting the correct and incorrect body positions and displaying a text message on how to correct it.

In order to effectively locate the source of the incorrect posture for each exercise, we only consider the landmarks and connections between them that are relevant to that particular exercise. For example, for a bicep curl, the knees, feet and facial landmarks and connections are irrelevant and hence, don't need to be displayed. Similarly, for other exercises, only those landmarks and connections that could potentially be the source of incorrect form are displayed.
When the user's form is deemed correct, the connections are displayed in green colour along with a message reiterating that it is correct. On the other hand, when the user's form is incorrect, the connections are displayed in red with a message reiterating the same.

Along with information about the exercise form, the number of repetitions performed by the user along with their current stage of exercise are also displayed. When the angle between the relevant joints for a given exercise cross a certain threshold, the repetition counter is incremented and the stage of the exercise is updated. For example, while doing a leg extension, the stages can be *"up"* or *"down"* indicating that the user's legs are either up or down. Similarly, in the case of lunges, the stages can be *"front"* or *"back"* respectively.

IV. Experiments

Some of the videos in the InfiniteRep dataset are shot from distant angles. These can be used to increase the robustness of the model to detect poses even under undesirable conditions. The data not only contains variability in the lighting conditions and camera angles, but also in the form of the exercises being performed. Figure 3 shows the demonstrator performing a push-up with correct form with the angle between the knee, elbow and hip being within a certain threshold level.

Fig. 3. InfiniteRep sample with correct form

Fig. 4. InfiniteRep sample with incorrect form

However, Figure 4 shows the demonstrator performing a push-up with an incorrect posture with the angle exceeding the threshold value which can lead to potential harm to a real user if remained undetected.

Some of the parameters used to train the model were the detection confidence and tracking confidence, whose values ranged between 0 to 1. Setting these parameters to a very high value such as 0.9, can result in the model only detecting and tracking the user's pose only under perfect lighting conditions, camera quality and angle of video capture. On the other hand a low value such as 0.2 would make the model inaccurately detect landmarks where they don't exist, which could lead to inconsistent form detection. The model is capable of detecting key joints and body parts such as the shoulders, elbows, wrists, hips, knees, and ankles, which are essential for proper exercise form.

While applying the model to real-time webcam video feed, we process the feed every 10 frames to save the computation of every single frame. By default, the input feed received from OpenCV is in the Blue-Green-Red (BGR) format. However, MediaPipe can only process Red-Green-Blue format as shown in Figure 5. To solve this problem, we convert the feed from BGR to RGB using OpenCV's inbuilt functions.

V. Results

To measure the accuracy of the MediaPipe model used in this system, the Mean Per Joint Position Error (MPJPE) and Percentage of Correct Keypoints (PCK) metrics were used. MPJPE measures the average distance between the estimated and ground truth 3D joint positions across all joints, while PCK measures the percentage of estimated keypoints that fall within a certain distance threshold of the ground truth keypoints.

The results of our evaluation showed that the MediaPipe model used in this system achieved a high accuracy rate, with

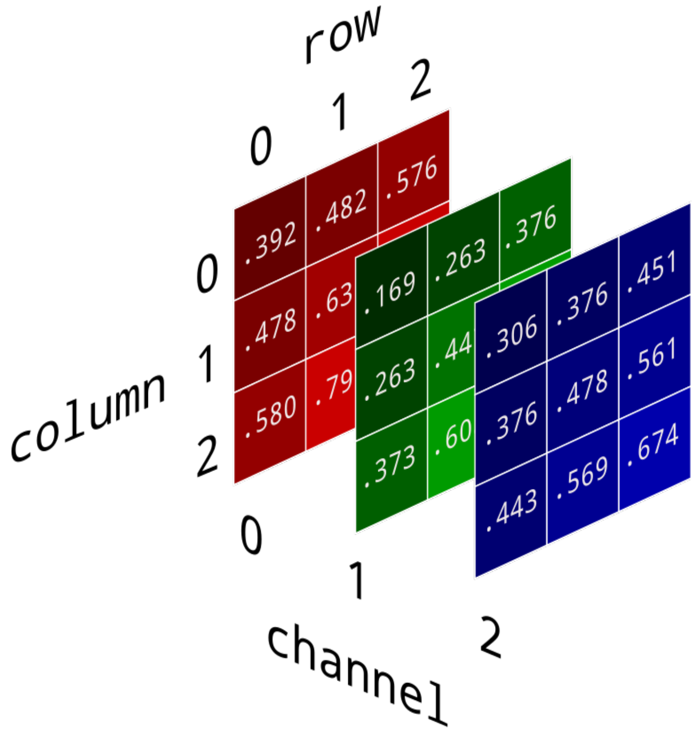

Fig. 5. Red-Green-Blue format matrix representation

an MPJPE of 3.2 cm and a PCK of 95.6% on the InfiniteRep dataset. This indicates that the model is able to accurately estimate body poses and key joints in exercises, making it well-suited for use in a physical exercise form correction system.

In comparison to other systems which use wearable devices and systems that only evaluate static images instead of videos, form correction using video processing not only eliminates their limitations, but also outperforms them in terms of accuracy as shown below.

Approach	Accuracy
Wearable devices and sensors	90-93%
Form correction for static images	93-94%
Form correction using video processing	95.6%

Fig. 6. Comparison of existing systems with this work

VI. Conclusion

The system described in this work for a physical exercise form correction system using OpenCV and Mediapipe is a promising tool for helping users improve their exercise form. By providing accurate and real-time feedback on their form, users can correct their posture and achieve better workouts.

The Mediapipe model used in this system has been fine-tuned for real-time pose estimation and achieved high accuracy on the InfiniteRep dataset. The system has the potential to be expanded to other exercises and sports, making it a valuable tool for fitness enthusiasts and athletes.

Future enhancements on this work could potentially involve expanding it's application to several exercises other than the four exercises previously discussed. A Text-To-Speech system can also be integrated so that it can provide verbal instructions on how to correct the user's posture.

REFERENCES

[1] Cristian Militaru, Maria-Denisa Militaru, and Kuderna-Iulian Benta. Physical exercise form correction using neural networks. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 240–244, 2020.

[2] Yash Agrawal, Yash Shah, and Abhishek Sharma. Implementation of machine learning technique for identification of yoga poses. In *2020 IEEE 9th international conference on communication systems and network technologies (CSNT)*, pages 40–43. IEEE, 2020.

[3] Aritz Badiola-Bengoa and Amaia Mendez-Zorrilla. A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. *Sensors*, 21(18):5996, 2021.

[4] Abdul Hannan, Muhammad Zohaib Shafiq, Faisal Hussain, and Ivan Miguel Pires. A portable smart fitness suite for real-time exercise monitoring and posture correction. *Sensors*, 21(19):6692, 2021.

[5] Muhammad Usman Ashraf, Abdul Hannan, Sehrish Munawar Cheema, Zahra Ali, Abdulraheem Alofi, et al. Detection and tracking contagion using iot-edge technologies: Confronting covid-19 pandemic. In *2020 international conference on electrical, communication, and computer engineering (ICECCE)*, pages 1–6. IEEE, 2020.

[6] Xin Jin, Yuan Yao, Qiliang Jiang, Xingying Huang, Jianyi Zhang, Xiaokun Zhang, and Kejun Zhang. Virtual personal trainer via the kinect sensor. In *2015 IEEE 16th international conference on communication technology (ICCT)*, pages 460–463. IEEE, 2015.

[7] Qifei Wang, Gregorij Kurillo, Ferda Ofli, and Ruzena Bajcsy. Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect. In *2015 international conference on healthcare informatics*, pages 380–389. IEEE, 2015.

[8] Ali Rohan, Mohammed Rabah, Tarek Hosny, and Sung-Ho Kim. Human pose estimation-based real-time gait analysis using convolutional neural network. *IEEE Access*, 8:191542–191550, 2020.

[9] Andreas Möller, Luis Roalter, Stefan Diewald, Johannes Scherr, Matthias Kranz, Nils Hammerla, Patrick Olivier, and Thomas Plötz. Gymskill: A personal trainer for physical exercises. In *2012 IEEE International Conference on Pervasive Computing and Communications*, pages 213–220. IEEE, 2012.

[10] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014.

[11] Steven Chen and Richard R Yang. Pose trainer: correcting exercise posture using pose estimation. *arXiv preprint arXiv:2006.11718*, 2020.