RepRight: AI Based Gym Form Correction App

Comparison Study on YoloV7 and MediaPipe as key body point extraction models

Moulya Shetty
Computer Science and Engineering
PES University
Bangalore, India
moulyashetty26@gmail.com

Nameeta Kuruwatti
Computer Science and Engineering
PES University
Bangalore, India
nameeta.kuruwatti@gmail.com

Prajay V K
Computer Science and Engineering
PES University
Bangalore, India
prajay827742@gmail.com

Nandita Pydikondala
Computer Science and Engineering
PES University
Bangalore, India
nanditapydi@gmail.com

Dr. Sandesh B J
Computer Science and Engineering
PES University
Bangalore, India
sandesh$_bj$@$pes.edu$

***Abstract*—In response to the burgeoning interest in a healthier lifestyle, our project focuses on the growing gym culture, specifically in the realm of weight training. There is a prevalent issue of individuals risking injury due to improper form or failing to optimize their workout outcomes. Recognizing the constraints posed by the cost and availability of personal trainers, our app harnesses the power of AI to address this problem. By offering real-time feedback to users, our app helps guide them in correcting their exercise forms and derive maximum benefits. The Gym Posture Recognition project aims to develop a machine learning model that can assess different postures during exercises in a gym or fitness setting. The model can be hosted on an app and a mobile phone camera can be used to capture video input of the user performing the exercise. The machine learning algorithm can then analyze the input and provide feedback on the user's form. The goal of this project is to help users improve their exercise techniques and reduce the risk of injury by providing accurate feedback on their posture based on a large dataset.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The gym industry, especially weight training, is becoming increasingly popular due to its many health benefits. However many individuals either injure themselves due to incorrect form or are unable to maximize the benefits of each exercise. We propose an innovative solution that utilizes machine learning to analyze and provide real-time feedback on users' exercise performances, all recorded through a simple smartphone camera. Human Pose estimation is a vital computer vision technique that is designed to recognize and track specific points on the human body. This benefits our use immensely. While building our project, several pose estimation models were explored. The leading pose estimation model, Google's MediaPipe, is known for its highly accurate body key point extraction. However, when building our project with it, we found peculiar faults. This led us to discover an equally accurate, newer model - the YoloV7 pose estimation model. This paper explores the key differences between the two models and breaks down their performance in varying scenarios of our use case. The findings of this study will provide valuable insights into the suitability of MediaPipe and YOLOv7 for different requirements.

Identify applicable funding agency here. If none, delete this.

II. Related Works

The paper presents an approach to improve gym exercise form of a user by assessing their pose. The pose estimation domain has been dominated by Mediapipe and only recent applications have applied YOLOv7. This section involves related works on Mediapipe with our current use case and applications of YOLOv7 for pose estimation.

A. Exercise Form Correction with Mediapipe

Existing approaches focusing on the Exercise Form correction domain commonly used Mediapipe and at times additional algorithms to extract the key points that make up the pose. The aforementioned algorithms include Dynamic Time Warping(DTW). This was used to sync user and trainer videos to obtain frame-pair mappings to compare. Another involved non-real-time processing of a recorded video which performed the classification by using a nearest neighbour classifier, however, this had the drawback of not being able to handle different angles which could be associated with mediapipe's performance. Moreover, a Stacked Hourglass neural network solution was implemented for pose estimation.

The paper "Pose Estimation..."[3] demonstrated a virtual fitness trainer with Google's Mediapipe library, which was made for various multimodal machine learning and deep learning pipelines. Mediapipe's deep algorithms and posture estimation module are utilized to develop this novel system, which captures user movements by identifying unique body landmarks for each exercise. The system determines the user's performance by computing angles and landmarks and feeding

this data into a machine-learning model. MediaPipe faces challenges when operating in densely populated environments.This may lead to misinterpretation of the body's position. Occluded body parts also pose a challenge for MediaPipe's pose calculation. It suggests using multi-person pose estimation to enhance pose estimation in crowded scenarios to improve accuracy.

Another method was developed to detect human deep squat with the help of mediapipe in the paper "Human deep squat..."[4]. The model constructed from the approach of combining the MediaPipe algorithm and the improved YOLOv5 network produced a faster model. The high detection speed and lightweight nature of it makes it ideal for apps and deployable on mobile phones.

Meanwhile, "Fitcercise"[5] used a novel approach of combining blazepose and MediaPipe to get the 33 key points for pose estimation. This was made possible using a double step teacher pipeline. BlazePose is an algorithm meant to extract key points efficiently on a mobile phone gpu which makes it appropriate for an app.It has currently only implemented the upper half of the body.

Another paper

B. YOLOv7 for Pose Estimation

The paper by Hung-Cuong Nguyen[1] involves building an application to support rehabilitation of the hand after surgery. It was concluded that accurate assessments can be made using computer vision, AI and deep learning. A model proposed to handle this involves hand-tracking detection before recognising the activity. This hand tracking has been made possible using different pose estimation models and the best results were found to be with YOLOv7. This paper sheds light on how well YOLOv7 performs however it focuses mainly on the keypoints involving the hands. The goal of our research focuses on other body parts excluding the fingers and the face.

Another paper by Henry O. Velesaca et al[2] was written with the goal of Human Height Estimation. They presented a two-step approach where the first step involved object detection, the human and the second step on top of object detection to calculate the height. Multiple methods to proceed with each of these steps were used. The second step involved different formulas which we are not concerned with for our research. However, the first stage of the height estimation through object detection compared YOLOv7 and Mediapipe which gave us insight into the statistical proof of the difference between the two in that particular scenario. This navigated us towards comparing accuracy of YOLOv7 an

III. Dataset

The dataset for our model contains approximately 900 images of individuals performing squats. This dataset was categorized into three distinct labels: "correct," "too low," and "too high." These labels were essential for providing users with precise feedback, enabling them to improve their squat form effectively.

The dataset contains images of people performing squats in various settings. The categories are as follows- body weight squats, Weighted squats, occluded person, low light conditions, low-quality images, high-quality images and multiple people in the frame.

IV. Methodology

The two pose estimation models, MediaPipe model and YoloV7 pose estimation model, will be tested on images of varying scenarios. Based upon the ability to identify key points and their accuracy, the models will be compared and checked to find the better-suited model.

A. Body weight squats

This category of the dataset contains images of people performing squats solely on body weight. This means the person is not holding any weight or dumbbell to obstruct the model from detecting all key body points in the image.

TABLE I
BODY WEIGHT SQUATS

Bodyweight Squats	**Model**	
	MediaPipe	***YoloV7***
Left Ankle	6	6
Right Ankle	6	6
Left Knee	6	6
Right Knee	5	6
Left Hip	5	6
Right Hip	3	6
Left Shoulder	5	6
Right Shoulder	4	6
Left Elbow	3	6
Right Elbow	3	6
Left Wrist	3	6
Right Wrist	4	6

Total images = 6

B. Weighted Squats

This category of the dataset contains images of people performing squats using weights. The weights could be in the form of a dumbbell or barbell. This leads to a change in the position of arms and exercise form compared to body weight squats. However, camera angles are taken where the person is not obstructed or covered by the weights they carry. Thus the model has visibility to detect all key body points.

C. Occluded Person

This category of the dataset contains images of people performing squats using a barbell or in a squat rack. This leads to parts of the person being blocked by the weight or gym equipment being used. Thus the model's performance and accuracy on the occluded parts of the person is tested.

D. Low Light Settings

This category of the dataset contains images of people performing squats in low light conditions or bad lighting conditions. This leads to low visibility for the model on which its performance and accuracy is tested.

TABLE II
WEIGHTED SQUATS

Key Body Points	Model	
	MediaPipe	***YoloV7***
Left Ankle	6	6
Right Ankle	6	6
Left Knee	6	6
Right Knee	5	6
Left Hip	5	6
Right Hip	3	6
Left Shoulder	5	6
Right Shoulder	4	6
Left Elbow	3	6
Right Elbow	3	6
Left Wrist	3	6
Right Wrist	4	6

Total images = 6

TABLE III
OCCLUDED PERSON

Key Body Points	Model	
	MediaPipe	***YoloV7***
Left Ankle	4	6
Right Ankle	3	4
Left Knee	4	6
Right Knee	2	4
Left Hip	3	4
Right Hip	5	4
Left Shoulder	4	6
Right Shoulder	4	5
Left Elbow	3	6
Right Elbow	3	3
Left Wrist	3	4
Right Wrist	1	3

Total images = 6

TABLE IV
LOW LIGHT SETTINGS

Key Body Points	Model	
	MediaPipe	***YoloV7***
Left Ankle	4	5
Right Ankle	5	5
Left Knee	4	5
Right Knee	4	5
Left Hip	5	4
Right Hip	4	4
Left Shoulder	5	5
Right Shoulder	4	5
Left Elbow	4	5
Right Elbow	5	4
Left Wrist	2	3
Right Wrist	2	3

Total images = 5

E. Low quality images

This category of the dataset contains images of people performing squats in low light conditions or bad lighting conditions. This leads to low visibility for the model on which its performance and accuracy is tested.

TABLE V
LOW LIGHT SETTING

Key Body Points	Model	
	MediaPipe	***YoloV7***
Left Ankle	count1	count2
Right Ankle	count3	count4
Left Knee	count1	count2
Right Knee	count3	count4
Left Hip	count1	count2
Right Hip	count3	count4
Left Shoulder	count1	count2
Right Shoulder	count3	count4
Left Elbow	count1	count2
Right Elbow	count3	count4
Left Wrist	count1	count2
Right Wrist	count3	count4

Total images = 6

F. Multiple people in frame

This category of the dataset contains images of people performing squats with people present in the background. This tests the model on accuracy and performance in identifying all the people in the frame or identifying only a single individual.

TABLE VI
TABLE TYPE STYLES

Key Body Points	Model	
	MediaPipe	***YoloV7***
Left Ankle	count1	count2
Right Ankle	count3	count4
Left Knee	count1	count2
Right Knee	count3	count4
Left Hip	count1	count2
Right Hip	count3	count4
Left Shoulder	count1	count2
Right Shoulder	count3	count4
Left Elbow	count1	count2
Right Elbow	count3	count4
Left Wrist	count1	count2
Right Wrist	count3	count4

Total images = 6

V. RESULTS

Results from the tests performed:

VI. DISCUSSION

Breaking down code with results

ACKNOWLEDGMENT

I would like to express my gratitude to Dr. Sandesh B J, Chairperson, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this UE20CS461A - Capstone Project Phase – 2. I am grateful to

the Capstone Project Coordinator, Dr. Sarasvathi V, Professor and Dr. Sudeepa Roy Dey, Associate Professor, for organizing, managing, and helping with the entire process. I take this opportunity to thank Dr. Sandesh B J, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help. I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University and Prof. Nagarjuna Sadineni, Pro-Vice Chancellor - PES University, for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement I have received from my family and friends.

REFERENCES

[1] [1] Yucheng Chen, Yingli Tian, Mingyi He,Monocular human pose estimation: A survey of deep learning-based methods,Computer Vision and Image Understanding,Volume 192, 2020, 102897, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2019.102897.

[2] [2] G. Taware, R. Kharat, P. Dhende, P. Jondhalekar and R. Agrawal, "AI-Based Workout Assistant and Fitness Guide," 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, Pune, India, 2022, pp. 1-4, doi:10.1109/ICCUBEA54992.2022.10010733.

[3] [3] A. Nagarkoti, R. Teotia, A. K. Mahale and P. K. Das, "Realtime Indoor Workout Analysis Using Machine Learning Computer Vision," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 1440-1443, doi: 10.1109/EMBC.2019.8856547.

[4] [4] Q. -T. Pham et al., "Automatic recognition and assessment of physical exercises from RGB images," 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE), Nha Trang, Vietnam, 2022, pp. 349-354, doi: 10.1109/ICCE55644.2022.9852094.

[5] [5] Pose Estimation and Correcting Exercise Posture: Rahul Ravikant Kanase, Akash Narayan Kumavat, Rohit Datta Sinalkar, Sakshi Somani ITM Web Conf. 40 03031 (2021) DOI: 10.1051/itmconf/20214003031

[6] [6] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 1297-1304, doi: 10.1109/CVPR.2011.5995316.

[7] [7] H. O. Velesaca, J. Vulgarin and B. X. Vintimilla, "Deep Learning-based Human Height Estimation from a Stereo Vision System," 2023 IEEE 13th International Conference on Pattern Recognition Systems (ICPRS), Guayaquil, Ecuador, 2023, pp. 1-7, doi: 10.1109/ICPRS58416.2023.10179079.

[8] [8]S. -H. Nguyen et al., "Segmentation and observation of hand rehabilitation exercises by supporting of acceleration signals," 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Taipei, Taiwan, 2023, pp. 1291-1295, doi: 10.1109/APSIPAASC58517.2023.1031730