Available online at http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 1, issue 10, pp. 240 - 245, Dec. -2021
https://doi.org/10.53633/ijmasri.2021.1.10.001

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

GYM REP TRACKER USING MEDIAPIPE AND PYTHON

[1]Rahul Ratusaria, [1]Tushar Baghel, [1]Ayush Chander Vanshi and [1] Neeraj Garg

[1] Computer Science and Engineering Maharaja Agrasen Institute of Technology, Rohini, Delhi

Abstract

Human Pose estimation has grabbed the eye of the computer vision community for the past few decades. It is a vital step closer to knowledge people in pics and motion pictures. Strong articulations, small and hardly visible joints, occlusions, apparel, and lighting changes make it very difficult to perform estimate pose. Human Pose estimation is an important problem that needed to be study. It is used to detect human anatomical key points (e.g., shoulder, elbows, legs, wrist, etc.) in real time using less computational resources. There are many Artificial Intelligence models i.e, Posenet, OpenPose[1] and MediaPipe[8] for Real time Human Pose Estimation. Many experiments has performed to find out the best suitable model for Human Pose Estimation. Experiments stated that PoseNet is suitable to run on lightweight devices like browsers whereas OpenPose meant to run on GPU powered devices and is more accurate. On the other hand, MediaPipe is very fast, modular, reusable and highly efficient. Hence, our model uses the MediaPipe to perform its estimation.

Keywords: Pose estimation, Gym Rep Tracker, Media Pipe, Python, Machine learning

Introduction

Human Pose estimation falls under the domain of Computer Vision which can be defined as the "a field of artificial intelligence in which computers are trained to interpret and understand the visuals of the world around us." It generates information that describes, interpret and explain input images or videos in a way as humans do and at the speed of thousands of images per second.

Human Pose estimation has enjoyed the attention of the Computer Vision community for the past few decades due to its increasingly large number of applications in the areas of human-machine interaction, virtual reality, elderly care, robotics, etc. It is a crucial step towards understanding people in images and videos. Human Pose Estimation defined as the problem of localization of human joints (also known as key points - elbows, wrists, etc.) in images or videos. It is also defined as the search for a specific pose in space of all articulated poses.

Clearly, improving technology of human pose estimation will enhance our knowledge of Computer Vision largely. Human pose estimation is used in various fields like fitness tracking, sign language recognition, gestural control, classification of images and recordings, activity acknowledgment, etc. With further success in this domain, Human Pose Estimation can also be used in active investigation of human PC interaction difficulty for the identification and localization of key points of the body that mainly includes various joints and body movement forecast and also shares difficulties in detection, for example in clustering, lighting, perspective, and scale the foreshortening of appendages, impediment of appendages, turn and introduction of the figure, and cover of different subjects. It can also be applied to other dense prediction tasks, e.g., semantic segmentation, object detection, face alignment, image translation, as well as the investigation on aggregating multiresolution representations in a less light way.

But the task of Human Pose Estimation, as of now, suffers from a variety of problems. To start with, when the pose estimation model is unsure about a pose, it still fully commits to its output and disregards uncertainty. This problem occurs due to motion blur as the network has difficulties to decide between left and right in this case. As pose estimation models have mostly seen forward facing persons these are generally more inclined towards predicting a forward-facing person in case of uncertainty. When left and right of a 2D prediction are incorrectly flipped in at least one of the views, the merged 3D prediction will collapse to the vertical line of the person resulting in poor 3D pose estimation.

Keeping its importance in mind, a lot of work has done on this domain but still, performance in predicting human poses needs improvement for real time used case scenarios. Considering the current situation of human pose estimation, we have chosen the MediaPipe for estimating the poses because of its easiness, fast and robust behaviour.

Our model is implemented for the fitness tracking purposes. In this, video of the person is captured in the real time using the Open CV framework of the python, after that various points of that person is estimated using Media Pipe and after that we will find the angles between the shoulder, elbow and wrist for estimating the up and down counter. In this sense, our model will work.

Related work

[4]Objectives to triumph over the issue ultra-modern detecting joints below occlusion for multi person pose estimation. The writer proposes Omni Pose, an unmarried-pass, quit-to-give up trainable framework that achieves state-of-art outcomes for multi-individual pose estimation. The use of a unique waterfall module, the Omni Pose structure leverages multi-scale characteristic representations that boom the effectiveness modern backbone characteristic extractors. Multi-scale representations effectively utilized in spine systems for pose estimation. Stacked hourglass[23] networks use cascaded systems of the hourglass approach for estimating poses.

[5]Proposes a simulation approach to detect distinct human poses in real time with streaming data. The proposed simulation mechanism describes the simulation of different poses of human and generates characteristic descriptors for every of the pose and trains the model the use of simple classifier. The skilled version predicts the real-time human pose detection on video streaming statistics. In the solution, the extraction of nodal factors of human skeleton based on pre-trained model the use of Deeper-cut algorithm[9]. The Deeper-cut set of rules based on ResNet structure wherein the model is educated with big MPII human pose information units and Coco records sets to extract nodal points or key points for you to be the human joints to shape the human

skeleton shape. Within the proposed model, the unique poses are expected with less frame charge using easy 2D cameras and with correct predictions by reducing the processing time and with less computational efforts. The answer is deployed in a cloud server Amazon internet carrier (AWS), and at the consumer side (personal laptop, laptop or mobile) the streaming application will be strolling.

[7]Provided an efficient framework for human poses estimation with two divisions, an efficient head with an efficient backbone. Using a differentiable neural architecture search method, he customized the backbone network design for pose estimation, and reduced computational cost degrading negligible accuracy. For the efficient head, narrowing the transposed convolutions and proposing spatial information correction module for promote the final prediction performance is the key. The network is evaluated on the COCO and MPII datasets.

[8]Is a light weight convolutional neural network architecture for estimating human pose that is tailored for real-time inference on mobile devices. The main contributions includes a novel body pose tracking solution and a lightweight body pose estimation neural network that uses both heatmaps and regression to key point coordinates. The significant speedup of the model with little to no quality degradation use an encoder-decoder network architecture to predict heatmaps for all joints, followed by another encoder that regresses directly to the coordinates of all joints. The key insight behind the work is that the heatmap branch can be discarded during inference, making it sufficiently lightweight to run on a mobile phone. Based on detecting face of person does not perform well when person is facing backwards. The model has 33 key points but it cannot recognize direction of fingers properly and performs worse than Open Pose[1].AR dataset.

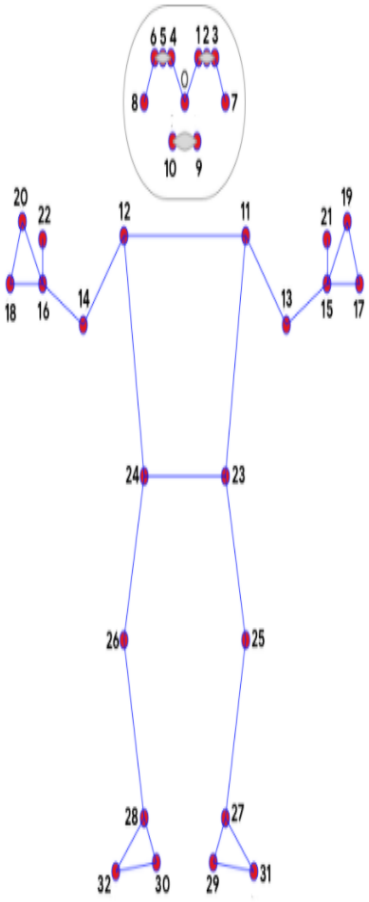
Methodology

Subsequently predicts the pose landmarks within the ROI using the ROI-cropped frame as input

Concept Used

MediaPipe Pose

MediaPipe Pose is a ML body pose tracking solution under high-fidelity, consisting 33 3D points overall body (or 25 upper-body landmarks) from RGB video frames and utilizes our Blaze Pose research powering the ML Kit Pose Detection API. Our method achieves real-time performance on most modern mobile phones, desktops/laptops, in python



and even on the web. Need of the hour is to implement our method in accordance with IOT.

A two-step detector-tracker ML pipeline is used, which proved to be effective in our MediaPipe Hands and MediaPipe Face Mesh solutions. The pipeline first locates the pose region-of-interest (ROI) within the frame using Detector. The tracker

MediaPipe consists of two landmark point versions: a full-body model predicting the location of 33 pose landmark points (see figure below), and an upper-body version predicting the first 25 points

which is more accurate to use when lower body parts are not in view. The model we have used predicts all the 33 points of the body.

Using the MediaPipe pose modules we are create only joint detectors which detects only hand joints and calculate the accuracy of performance if the palm joints are going at same level of shoulders then it will count one set and move further. The equation to find angle and showing accuracy is:

radians = np.arctan2(c[1]-b[1], c[0]-b[0]) - np.arctan2(a[1]-b[1], a[0]-b[0])

angle = np.abs(radians*180.0/np.pi)

where, a, b, and c represents the shoulder point, elbow point and wrist point respectively.

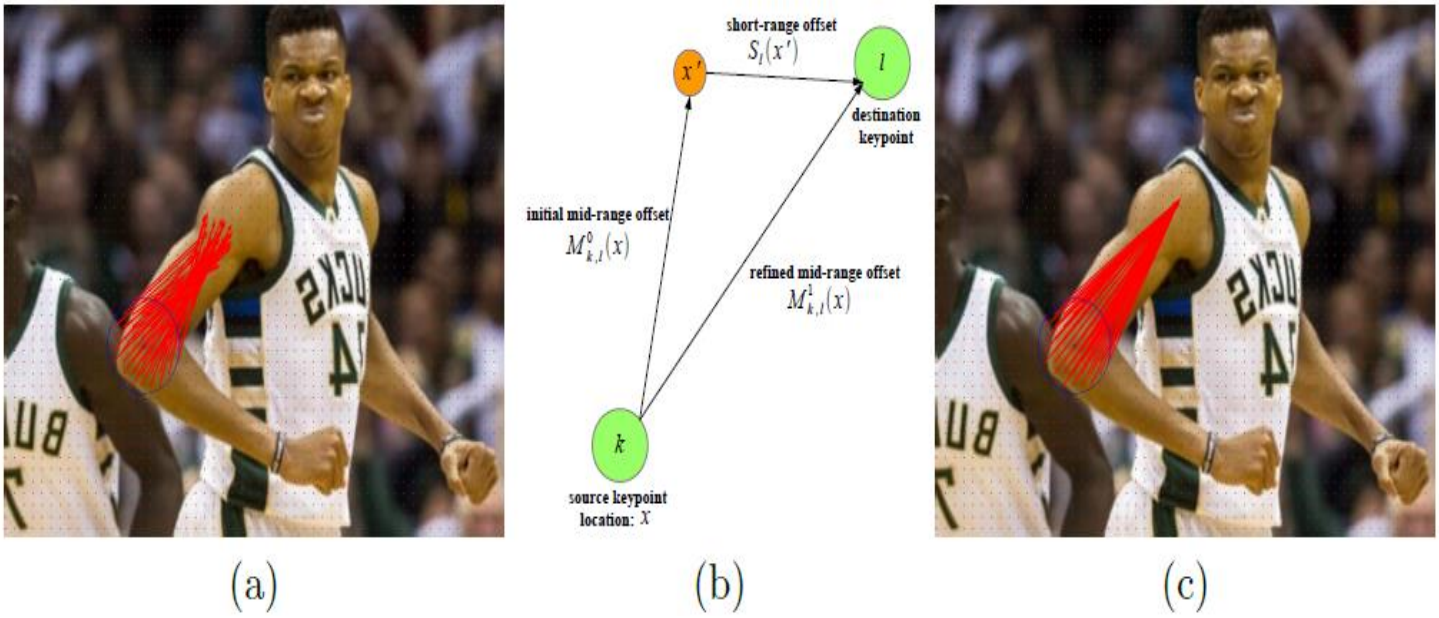


Fig. (C) [10] Mid-range offsets. (a) Initial mid-range offsets that starting around the Right Elbow keypoint, they point towards the Right Shoulder keypoint. (b) Mid-range offset refinement using the short-range offsets. (c) Mid-range offsets after refinements for each keypoint type.

Procedure

- Firstly, Install MediaPipe and OpenCV
- Than Import these two in the project.
- Than we have started making detections of images using webcam.
- After those angles between shoulder, elbow and wrist calculated using formula derived in Media Pose of 4.3 section of the article.
- At the very last, according to the conditions like if angle is greater than 160 degree, consider it as one down counter and if angle is less than 30 degree, consider it as one up counter. We have implemented the counter.

Evaluation metrics

- ***Percentage of Correct Key-points - PCK***

PCK calculates the percentage of detections that fall within a normalized distance of the ground truth. Detected joint is considered correct if the distance between the predicted and the true joint is within a certain threshold.

PCK@0.2 == Distance between predicted and true joint < 0.2 * torso diameter

The Percent of Correct Points with 20% tolerance (PCK@0.2) is used as evaluation metric.

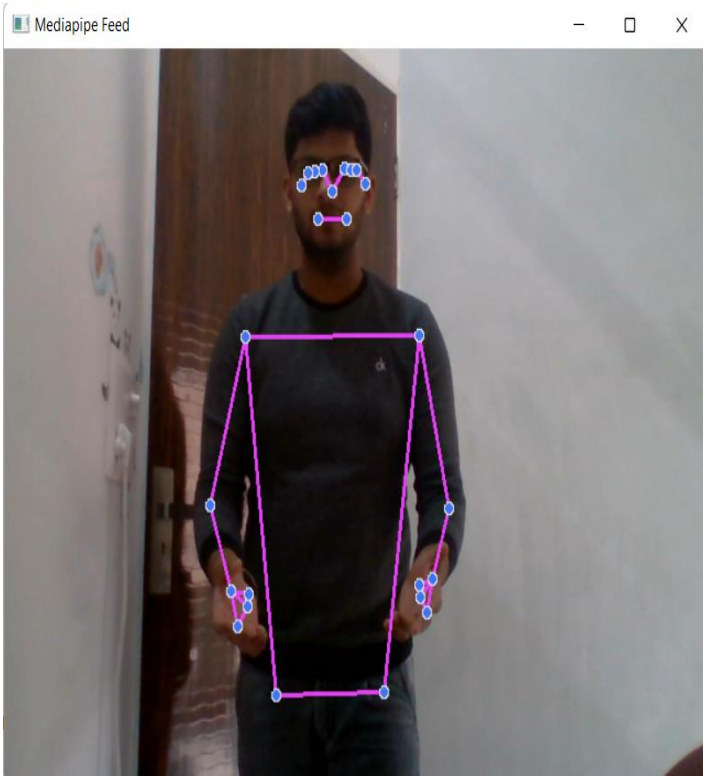
- ***Average Precision***

The concept of the average precision evaluation metric is mainly related to the competitive dataset. Basically, we need to consider any prediction with an IoU of more than or equal to 0.5 as a true positive.

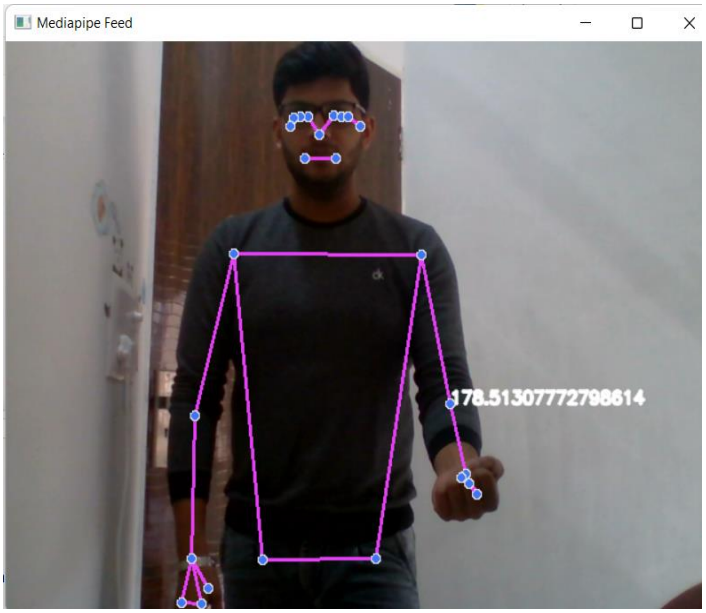
Result

After successfully implementing the Gym Rep counter, we have run the software and derived the results as follows:

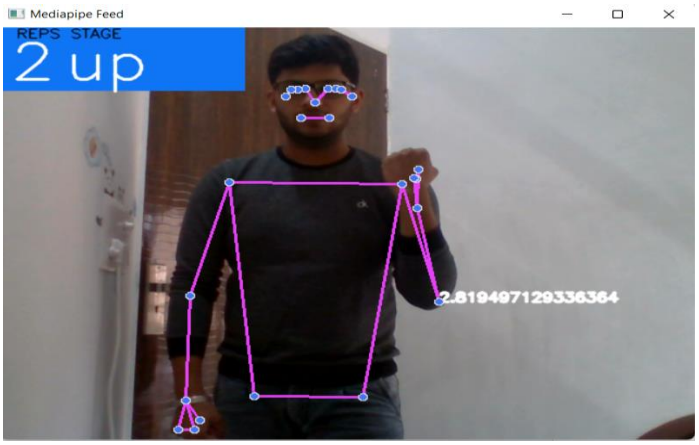
After estimating the joint coordinates, we have plotted those on the screen, which can be resulted as Fig (E):

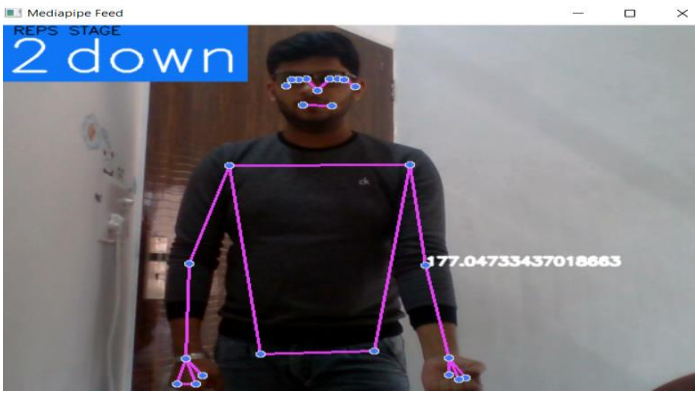


Than angle between shoulder, elbow and wrist is calculated and we tried to show that angle on the screen and result is as follows Fig. (F).



Till now we have implemented everything, we just now implemented the condition that if angle is greater than 160 than down counter will be shown and If angle is less than 30 than up counter will be shown and on completing 1 up and 1 down counter, counter will also be incremented.





Conclusion

In this paper, we implemented the Gym Rep Counter successfully using MediaPipe and Python. It is quite useful application especially during coronavirus pandemic because people and getting very lazy and do not have proper motivation to look after their body. So, this is developed to keep them tracked of their exercises performed.

Reference

1. Zhe Cao, 2019. Student Member, IEEE, Gines Hidalgo, Student Member, IEEE, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in CVPR.
2. George Papandreou., Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, and Kevin Murphy. 2018. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In The European Conference on Computer Vision (ECCV).
3. Valentin Bazarevsky., Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, Matthias Grundmann. 2020. BlazePose: On-device Real-time Body Pose tracking.
4. Bruno Artacho, Andreas Savakis. 2020. "OmniPose: A Multi-Scale Framework for Multi Person Pose Estimation". 2021
5. Alejandro Newell., Kaiyu Yang, and Jia Deng. 2016. Stacked hourglass networks for human pose estimation. In European Conference on Computer Vision (ECCV), pages 483–499.
6. Prasanth Kumar Ponnarassery, Gopichand Agnihotram, Pandurang Naik. 2020. Human Pose Simulation and Detection in Real Time Using Video Streaming Data.
7. Julian Tanke, Juergen Gall. 2021. Iterative Greedy Matching for 3D Human Pose Tracking from Multiple Views ,2021.
8. Wu Z, Shen C, Van Den Hengel A. 2019. Wider or deeper: revisiting the resnet model for visual recognition. Pattern Recogn.90:119–33.
9. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Lawrence Zitnick C. 2014. Microsoft coco: Common objects in context. In:European conference on computer vision; Cham:p. 740–755.
10. Wenqiang Zhang, Jiemin Fang, Xinggang Wang, Wenyu Liu. 2020. EfficientPose: Efficient Human Pose Estimation with Neural Architecture Search.

*****