www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 326-331

e-ISSN : 2583-1062
Impact Factor : 2.205

# SELF WORKOUT TRAINER

**Sanjay Srivastava[*1], Mayank Kushwaha[*2], Aditya Narayan Singh[*3], Ayush Tomar[*4], Mayank Pratap Puri[*5]**

[*1,2,3,4,5]Raj kumar Institute of Technology, Ghaziabad, Uttar Pradesh, India.

## ABSTRACT

Human pose estimation refers back to the technique of inferring poses in an photograph. Essentially, it entails predicting the posture of a person in vision. This trouble is likewise sometimes called the localization of human joints. It`s additionally critical to notice that pose estimation has diverse sub-duties which includes unmarried pose estimation, estimating poses in an photograph with many human beings, estimating poses in crowded places, and estimating poses in videos. It may be used in diverse extraordinary fields however our aim is to apply this records of joints from pose estimation to analyze extraordinary sporting events.

**Keywords:** Exercise, Fitness, artificial intelligence, Convolutional Neural Network(CNN), Dynamic Time Wrapping, Computer Vision  Microprocessor, Enginehead, L293D Current Amplifier, IRF 3205 MOSFET.

## 1. INTRODUCTION

Exercise may be very useful to the fitness if accomplished efficaciously however additionally may be very risky if posture and shape isn't accurate. Some sporting events which includes weight can purpose large twist of fate if shape is now no longer accurate. Our concept for this assignment is to broaden a software program or cellular app that may identify whether or not the man or woman is appearing workout efficaciously or now no longer and provide a few comments on how to accurate shape. There`s continually extra variety of human beings with inside the fitness center with appreciate to the variety of running shoes, so the teacher can`t preserve an eye fixed on everyone. That may purpose issues for the ones folks who're out of attain of a teacher at that moment. Our software program,Self Workout Trainer, will try to reduce that trouble with the aid of using tracking the ones human beings and giving them right comments. This will assist to lessen the injuries due to wrong postures while exercising.

First a part of Self Workout Trainer is Human Pose Estimation, which may be very tough however highly relevant subject in Computer Vision. Pose estimation is crucial for issues concerning human detection and hobby recognition, and also can useful resource in fixing complicated issues concerning human motion and posture.

Second Part of Self Workout Trainer is to recognize the exercise being performed and giving appropriate feedback on the workout they're doing.

Our plan is to have a whole software program which includes ideas that have been previously stated with characteristic of recording video to reveal and preserve music of user`s workout routine. It can even have characteristic of counting numbers of accurate reps and units and also will report user development on workout.

## 2. OBJECTIVES

The primary objective of this project are human pose estimation and posture correction from video.

Giving appropriate feedback to people and users during training must also be done.

So the listed milestone goals of this project are:

- Create Skeletal structure of person using body joints as keypoints.
- Near real-time pose recognition
- Recognize the type of the exercise the person is doing
- Provide positive or negative feedback to users based on their posture
- Develop a interface for user interaction

## 3. APPLICATION

This project focuses primarily on estimating the pose of the human body. Many for pose estimation Applications such as robot high-precision human recognition and intelligent driver support Self-driving car for pedestrian detection, human character animation, For medical applications such as medical training, detection of postural problems such as bad angle and bad grip. Our project focuses on postural correction in real-time. Take points from pose estimates and analyze them using different techniques, implement CNN model to know if the exercise performed is correct.

Skeletal Diagram Generated To Train Model

## 4. RELATED LITERATURE

While researching for this project, we found out that many research works have already been conducted in the field of human pose estimation but rarely any in the field of real-time posture correction for physical exercises.

Various researchers around the globe have used various methods to complete there objective, which is very informative.

Alexander Toshev and Christian Szegedy used Deep Neural Network method for human pose estimation[1]. They completed posture recognition using a DNN-based regression method based on body joints.

Novel architecture for Human posture estimation which has been proposed by researchers from New York University also has a major impact.. This method included 'position refinement' model which was trained to estimate the joint offset location in small parts of the image. This refinement model is jointly trained in cascade with a state-of-the-art ConvNet model to achieve improved accuracy

in human joint location estimation.[2].

Human Pose Estimation using Iterative error feedback method[3] has also been proposed.

Here they proposed a framework that expands the expressive power of hierarchical feature extractors to encompass both input and output spaces, by introducing top-down feedback.

Similar research has been done by using convolutional pose machines [4].Pose Machines provide a sequential prediction framework for learning rich implicit spatial models. In this work they showed a systematic design for how convolutional networks can be incorporated into the pose machine framework for learning image features and image-dependent spatial models for the task of pose estimation.

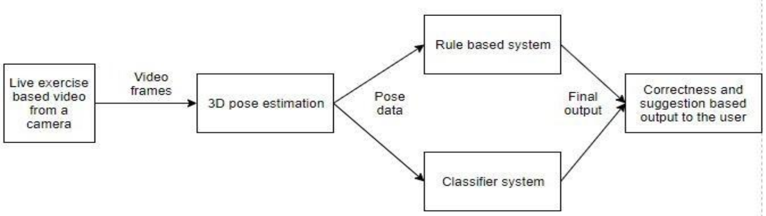
Similarly, there have been many researches in the field of exercise posture correction. Recently, Richard Yang and Steven Chan in 2018 proposed a PoseTrainer Model to correct exercise posture using PoseNet. They used Heuristic Based system to detect quality of exercise.

**DeepPose**

DeepPose was the first major paper[1], published in CVPR 2014 that applied Deep Learning to Human pose estimation. It achieved SOTA performance and beat existing models back in the year 2014. The model has an AlexNet backend and estimated pose in a holistic fashion, i.e certain poses can be estimated even if some joints are hidden when the pose is reasoned holistically. The paper applies Deep Learning (CNN) to pose estimation and kicked off research in this direction. The model used regression for XY locations for certain regions. This added complexity and weakened generalization hence performing poorly.

## 5. METHODOLOGY

### 5.1 System Design



www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 02, Issue 05, May 2022, pp : 326-331

e-ISSN : 2583-1062

Impact Factor : 2.205

Our software can work on all types of camera feeds, including mobile phones camera or webcam from a laptop computer. The camera is used to capture the standard RGB format video where the frame is fed to the 3D pose estimation model. Make sure that the person performing the exercise routine is clearly visible. Pose estimation is done to get detailed posture data, which is also sent to the rule-based system. As a classification and comparison-based system for analyzing workout routines. If done correctly, the number of repeated exercises will be automatically recorded and displayed in the application user interface. When a person does not exercise properly, based on the rules, feedback will be generated to notify the user and provide appropriate feedback.

### 5.2 Background Removal

By detecting changes in image sequence and segmenting the background we can facilitate easier object detection in foreground as it reduces noise. This is an important step in computer vision based image processing.

### 5.3 Posture Estimation

We start by recognizing humans in the image and draw a rectangle box around them this is called bounding box. Then comes the turn for joint locations.

One way to predict the skeletal structure is to create a confidence map for each joint location. The confidence map is a probability distribution over the image, representing the confidence value of common position for each pixel. There are many approaches to posture estimation and most of the popular methods involve deep learning techniques.

**OpenPose** which is currently the most popular bottom-up approaches for multi-person human posture estimation comes into use here. OpenPose first detects parts (keypoints) belonging to every person in the image, followed by assigning parts to distinct individuals. The OpenPose Network first extracts features from an image using first layers (VGG 19 in the diagram above). The features are then fed into two parallel branches of the convolution classes. The first branch predicts a set of 18 trust cards, each representing a specific part of the human postural skeleton. The second branch predicts a set of

38 affinity parts The field (PAF) represents the degree of association between the parties. The next steps are used to refine the predictions made by each branch. Use the coin's credit card, bipartisan graphs are formed between pairs of parts (as shown in the image above). Using the PAF values, The weakest links in the bipartite graph are removed. Through the above steps, the human postural skeleton can estimated and assigned to each person in the figure.

### VNECT:CNN Pose Regression

Our solution is based on a CNN that predicts in real- time both 2D and root (pelvis) relative 3D joint locations. In terms of 3D joint position accuracy, the new proposed fully-convolutional pose formulation achieves results comparable to state-of-the-art offline approaches. It can work without tight cropping around the topic because it is fully convolutional. Regardless of the scene conditions, the CNN can anticipate joint positions for a wide range of activities, giving a solid foundation for future pose refinement to create temporally consistent full-3D pose parameters.

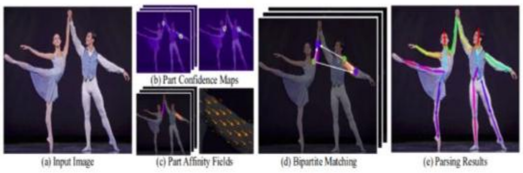



Figure 4: Vnect

### Kinematic Skeleton Fitting:

Using the CNN's 3D predictions and previous frames, a consistent full 3D skeletal pose can be generated. To create a smooth transition between frames, smoothing and filtering can be used. The bounding hierarchy can be transferred to the armature generated from the CNN output. The positions of the several armatures acquired from the CNN are adjusted relative to one another. This will make it easier to build up the animation system later. In addition, the size of the mesh we want to animate may vary; this can be addressed by tracking the pose in real metric space and then scaling it to suit the appropriate mesh.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 326-331

e-ISSN : 2583-1062

Impact Factor : 2.205

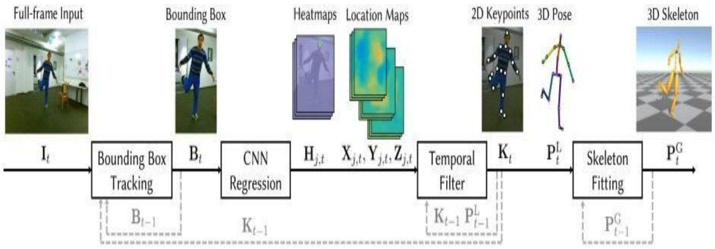



Figure 5: Kinematic Skeleton Fitting

## Exercise posture evaluation using a rule- based system:

To evaluate exercise posture, the output of pose estimate must first be processed into the right format. The person's entire skeleton estimation is collected, and any necessary normalisation is carried out. The torso length can be used as a benchmark for normalisation. The user's viewpoint is also determined.

The rule-based system will operate in accordance with the exercise's normal rules.

Heuristics and thresholds based on geometry can be used to determine whether or not an activity has been completed correctly. In addition, the user's genuine mistake might be highlighted. In a bicep curl, for example, we can find two heuristics of interest. To begin, the upper arm should be kept still and not move much. The angle between the upper arm vector and the torso vector can be used to quantify this. If the upper arm is held constant throughout the movie, it should be parallel to the body with slight fluctuations. Second, a correct, complete curl necessitates bringing the weight up until the bicep is fully contracted, which is usually done at 90 degrees between the upper arm and forearm.

This incorrect shape is usually caused by the user employing too heavy weights. The minimum angle reached between the upper arm and forearm can be used to measure this issue. The angle should be almost 180 degrees in the starting position with the weight down. The angle should decrease as the weight is hoisted until the user stops, then increase as the weight is brought down. Many additional exercises can benefit from such rules.

### 5.4 Evaluation based on Machine Learning

Using multiple machine learning approaches to assess the exercise posture is another option.

## Dynamic Time Warping (DTW):

Because the recorded films might be any length, each example will have a different keypoint vector length. Many machine learning models struggle with different feature vector lengths, therefore dynamic temporal warping (DTW) can help. The non-linear similarity of two time series is measured using the DTW metric. Because it is a direct comparison of two points at the same time, metrics like the Euclidean distance fail when two similar sequences are phase-shifted (i.e. altered in the time dimension). We may dynamically identify the keypoint in the second sequence that corresponds to a particular point in the first sequence using DTW. As a result, DTW distances may be determined and closest neighbour classification can be used to determine if an exercise sequence is proper or erroneous.

## Recurrent Neural Networks(RNN) and Long Short Term Memory (LSTM):

Recurrent Networks are preferred for sequence processing because they can remember the past. Recurrent Networks are made up of RNN cells. The input and the prior state are the two incoming connections in RNN cells. They have two outgoing connections as well, the output and the current state. This stage allows them to blend data from the past with present input.

A simple RNN cell is too basic to be employed consistently across domains for time series analysis. So many variants to adapt Recurrent Networks to diverse domains have been proposed throughout the years. LSTMs and GRUs are among them.

These can be employed in problems with time series. The drawback with this strategy is that for increased accuracy, a considerable amount of data may be necessary.

## Convolutional Neural Networks (CNN) - LSTMs:

RNNs can only remember the most recent past because state information goes through each timestep. Gated networks, such as LSTMs and GRUs, on the other hand, can accommodate longer sequences, but even they have their limits. We can learn more about this problem by looking at disappearing and exploding gradients, which occur in very deep networks. As a result, for extended sequences, we must shorten them. One method is to discard the signal's fine-grained time information. This can be accomplished by combining small groups of data points and constructing

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 326-331
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 2.205

features from them, which are then passed to the LSTM as a single datapoint. The data points can be sent to the recurrent units when CNN completes this process.

### 5.5 SYSTEM IMPLIMENTATION

### TensorFlow:

TensorFlow is a free open source dataflow and differentiable python library for a variety of tasks. This is an iconic library used in Machine learning applications such as neural networks.

### Tensorflow Lite:

This is used when your project contains an application for mobile devices. TensorFlow Lite is an open source deep learning framework for inference on devices. It forms the base to deploy machine learning models on mobile and IoT devices.

### OpenCV :

OpenCV (Open Source Computer vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state of the art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc.

## 6. INPUT MODALITY

Modality refers to the different types of inputs available. The input video/image may have different parameters. The modes of input can be as follows:

### 6.1 RGB image/video:

The images that we see around us on a daily basis, and the most common type of input for Pose Estimation. Models working on RGB-only input have a huge advantage over others in terms of the mobility of the input source. This is due to the ease of availability of common cameras(which capture RGB images), making them the models that can be used across a huge number of devices.

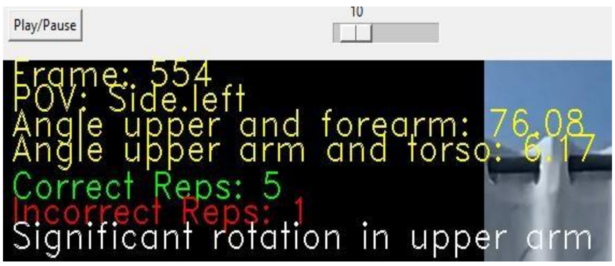
### 6.2 Depth information included image/video:

In a Depth image, the value of a pixels relates to the distance from the camera as measured by time-of- flight. The introduction and popularity of low-cost devices like Microsoft Kinect has made it easier to obtain Depth data. Depth image can complement RGB image to create more complex and accurate Computer Vision models, whereas Depth-only models are vastly used where privacy is a concern.

### 6.3 Infra-red (IR) image/video:

In an IR image, the value of a pixel is determined by the amount of infrared light reflected back to the camera. Experimentation in Computer Vision based on IR images are minimal, as compared to RGB and Depth images. Microsoft Kinect also provides IR image while recording. However, currently there are no datasets that contain IR images. IR based systems can have several other disadvantages as well.

## 7. RESULTING OUTCOME



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 326-331
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 2.205


Incorrect Rep Counted and feedback displayed Correct rep counted

## 8. FUTURE POSSIBILITIES

The application supports recognition and posture analysis for bicep-curls on recorded videos. The reps are counted correctly and feedback system is also suitable. In near future, we suggest few objectives:

1. Train models for front-raise and shoulder-press separately to include those exercise also.
2. Design web application to make this serviceavailable through cloud.
3. Redesign UI for the application and add button to switch between real-time camera feed and recorded videos.

## 9. CONCLUSION

During the course of this project our team worked with different python libraries to explore the possibilities of accomplishing our objective in innovative method. Our project is complete and our program successfully evaluates exercise posture while doing bicep-curls. There are much possibilities for future including the ones we suggested. The program evaluates exercise posture in almost real- time and is displayed using a basic UI.

## 10. REFERENCES

[1] D.Berndt and J. Clfford. Using dynamic time wraping to find patterns in time series. In AAAIWS, 1994.

[2] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. Black. Keep it simple: Autoatic estimation of 3d Human pose and shape form a single image. In ECCV,2016.

[3] Steven Chan and Richard yang, Research Paper on Pose Trainer: Correcting Exercise Posture using Pose estimation.

[4] H. Bristow, J. Valmadre, and S. Lucey. Dense semantic cor- respondence where every pixel is a classifier. In ICCV, 2015

[5] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille. Detect what you can: Detecting and representing objects using holistic models and body parts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1971–1978, 2014