International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 167-173
Published Online May 2020 in IJEAST (http://www.ijeast.com)

Auto_Fit: WORKOUT TRACKING USING POSE-ESTIMATION AND DNN

Nitesh Sonwani
UG Student
Department of CSE
Prestige Institute of Engineering
Indore, Madhya Pradesh, India

Aryan Pegwar
UG Student
Department of CSE
RNS Institute of Technology
Bangalore, Karnataka, India

Abstract— **Lack of physical fitness increases the risk of adverse health conditions including coronary heart diseases, high blood pressure, stroke, metabolic syndrome, type 2 diabetes which leads to a decrease in the life expectancy of humans. In our work, we have introduced Auto_fit, an application that suggests the workouts and tracks it. Auto_fit uses Postnet for doing pose estimation to find 17 body keypoints followed by using the DNN classifier to identify the state of exercise and then counts the repetitions performed. We collected the videos of trained professionals performing the exercise and then used it to train Auto_fit. Auto_fit takes live video feed and counts the repetitions of exercise performed. It works on two common exercises and can also be run on low single-board computers like Raspberry pi. Auto_fit helps in improving physical fitness and thus enables a person to live a longer and healthier life.**

Keywords— **DNN, Posenet, Raspberry pi, Pose estimation, computer vision.**

I. INTRODUCTION

Lack of physical fitness increases the risk of adverse health conditions including coronary heart diseases, high blood pressure, stroke, metabolic syndrome, type 2 diabetes which leads to a decrease in the life expectancy of humans.

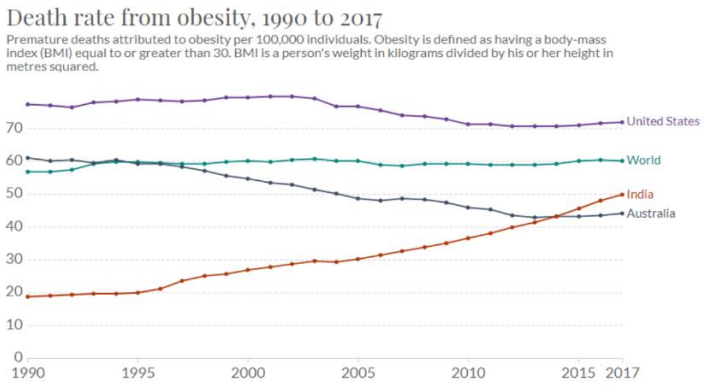

Fig 1. The death rate from obesity

From figure 1 we can conclude that the death rate due to obesity in India is less as compared to the world average but it's consistent growth is a reason for concern. Around 80% of adolescents are not physically active. Scarce of physical activity is a key factor for noncommunicable diseases like cardiovascular diseases, cancer, diabetes, and many more. In the U.S. more than 80% of adults and adolescents do not meet the guidelines for the physical activity mentioned by the Department of Health & Human Services. It is seen that urban, richer, and middle-aged population is more prone to be obese in India. Women who aged around 30 are more likely to be overweight as compared to men at the same age. This is happening due to the social customs which confines the agility and physical activity for women.

This is due to the sedentary lifestyle in the modern world. Physical fitness can be improved by sticking to a regular workout routine. And significant results can be seen by only doing basic workouts. Many people are also willing to start but due to a lack of knowledge of exercises and proper guidance, they are unable to inculcate this in their daily routine. And due to busy schedules, they are also unable to go to gyms or fitness centers to get proper guidance for their workout. In this paper, we are proposing a system that suggests the correct form of exercise to the person and also keeps track of the exercises performed by an individual using pose estimation(action recognition). This enables the person to indulge in physical activities at their own pace of time without having dependencies on others.

Auto_fit is particularly helpful for those people who can not or don't want to go out of their house for workouts. Auto_fit provides a complete workout plan for an individual which involves different types of activities which are divided into 6 days of a week. Auto_fit also checks for correct posture and notify if the posture is wrong. It also records the count of sets and repetitions performed for every exercise and keeps records of all of them to provide better suggestions for workouts in subsequent weeks. It can log and track data of at most 10 people at a time.

In the first part, it creates the skeleton of the user with 17 points like wrist, knees, ankles, etc. these points are

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 167-173
Published Online May 2020 in IJEAST (http://www.ijeast.com)

identified using the TensorFlow Postnet model which uses Mobilenet_V1 under the hood. And in the second part of the application, The identified points are then passed to a deep neural network model which is trained to identify the correct pose and count the repetitions of an exercise through live camera feed. Both parts of the application are combined to provide a complete solution for doing workouts at home.
This application is designed in such a way that it can be run on low powered computing devices like Raspberry Pi, Jetson nano, etc. to provide a compact and low-cost device that can be easily installed at home.

II. PROBLEM STATEMENT

The problem consists of identifying the 17 body key points in a video frame and then classifying the pose by giving input to a deep neural network that identifies the action performed by the person in that frame. The purpose is to train a deep neural network to identify the correct forms of the exercise.

III. RELATED WORK

[2][3] We have Identified various techniques that are used for human pose recognition. Different sensors and devices are used for pose estimation. In a paper, Reimers et al (2012) [1] have used open pose for recognizing some poses. [9] uses a stacked hourglass neural network architecture that in turn uses repeated bottom-up and top-down processing to achieve single pose predictions. In this Muhammad Usama et al (2017) [4] have used Microsoft Kinect to get real-time human joints points and using them to identify the correct yoga pose of a person. Shih-En et al (2016) [5] proposes a different architecture that uses multiple convolutional networks to improve the joint estimates over sequential passes. SVM technique is used to classify body posture from a 3D visual hull constructed from the input data source in [6] which as output give the recognized pose in the form of the thumbnail image Bogo et al (2016)[7] estimate the 3D pose and 3D mesh shape by using only a single RGB image.

Various computer vision-based systems have focused on the restoration of a healthy life. More recently, systems such as OpenPose [8,9], V NECT [10], and Adversarial PostNet. [4] have employed deep learning-based approaches to track pose from an RGB feed. Nevertheless, such procedures typically bank on estimating the skeleton of the user; this technique can be undependable when users are at a greater distance from the cameras and there is substantial occlusion, this scenario is very likely to happen at the gym. As far as we have discovered there was no former computer vision-based exercise recognition and tracking system has been established.

In [11] the author has used convolutional neural networks for doing image classification with TensorFlow.Paper [12] has formulated the idea of a spectral-spatial feature learning (SSFL) method to gather important features of hyperspectral images (HSIs). LeCun et al (2015) [13] describes useful concepts for deep supervised learning, unsupervised learning, reinforcement learning & evolutionary computation, and indirect search for short programs encoding deep and large networks.

IV. TECHNICAL APPROACH

Pipeline overview, talking about Auto_fit's technical aspects consists of two phases training and testing phase in which each phase is divided into multiple stages. Which are shown in figure 2 and figure 3. Here the process starts from giving input for training and getting output for the identified pose.

1.Training Phase

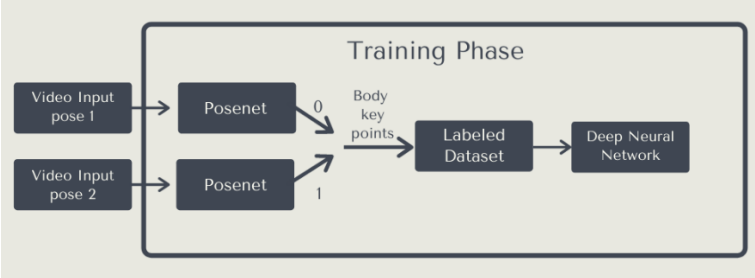

Fig 2 Training Phase of Deep Neural Network.

In this phase first, we take the exercise and divide it into initial and final states. Then we divide video into two parts and label it as initial and final states. Then both the videos are feed into the posenet model which generates the coordinates of body key points for corresponding videos. Then those coordinates are combined into a single dataset and are given as input to the DNN classifier for training.
Training is done until more than 90% accuracy is achieved. After training the DNN classifier, the model and it's weights have been saved. This process is done for one exercise and is to be repeated for every exercise.

2. Testing Phase

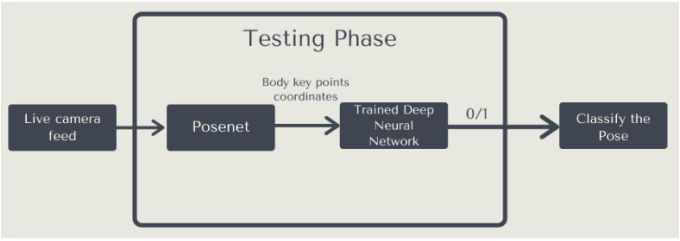

Fig 3. Testing Phase of Deep Neural Network

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 167-173
Published Online May 2020 in IJEAST (http://www.ijeast.com)

For testing, the video stream is taken live from the camera and fed into the posenet model which generates the coordinates of body key-points in that frame. Those keypoint coordinates are then fed into the DNN classifier which we have trained in the first phase. Then the DNN classifier identifies the pose as the initial or final state of the exercise. This output is used for further calculations like counting the reparations of an exercise.

Video Recording for the training dataset
We have first taken video of trained professionals performing a selected exercise, then we have cropped and trimmed the video so that sections only include the exercise. Then we divided the video into 2 separate videos which include the initial and final states of the exercise. These videos are used for training the DNN classifier.
The video is selected such that all body parts are clearly visible.

Live camera feed
There are no specific requirements for camera type or distance from the camera but the user needs to make sure of the following things:-

1. The body should be clearly visible in the frame
2. Ample amount of light is available
3. The video should not be blurry.

V. POSE ESTIMATOR

[14] Pose estimation refers to computer vision techniques that detect human figures in images and videos, so that one could determine, for example, where someone's elbow shows up in an image. For pose estimation, we have used deep convolutional neural networks (CNNs) to form pose coordinates. After experimentation with multiple state-of-the-art pose estimators, we choose to use the pre-trained model, posenet, for pose detection. PoseNet is a vision model that is used to estimate the pose of a person in an image or video by estimating where key body joints are.
[15] The PoseNet model is image size invariant, which means it can predict pose positions on the same scale as the original image regardless of whether the image is downscaled. This means PoseNet can be configured to have a higher accuracy at the expense of performance. Performance varies based on your device and output stride (heatmaps and offset vectors).
There are two main approaches to doing pose estimation:-

I. *The top-down approach* starts by identifying and roughly localizing person instances using a bounding box object detector, followed by single-person pose estimation or binary foreground/ background segmentation in the region inside the bounding box
II. *The bottom-up approach* starts by localizing identity-free semantic entities (individual keypoint proposals or semantic person segmentation labels, respectively and then grouping them into person instances.

The following image is passed on to the model and the result is generated with marked key points on the image.

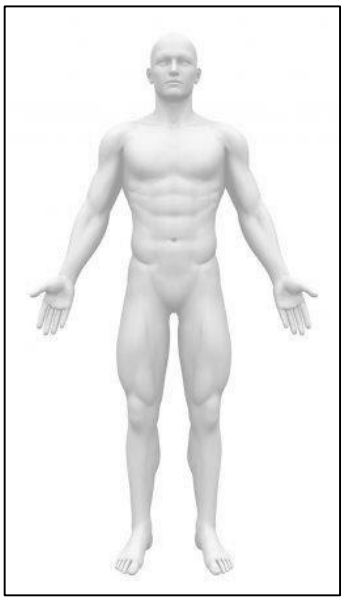

Fig 4. Input Image to Posenet

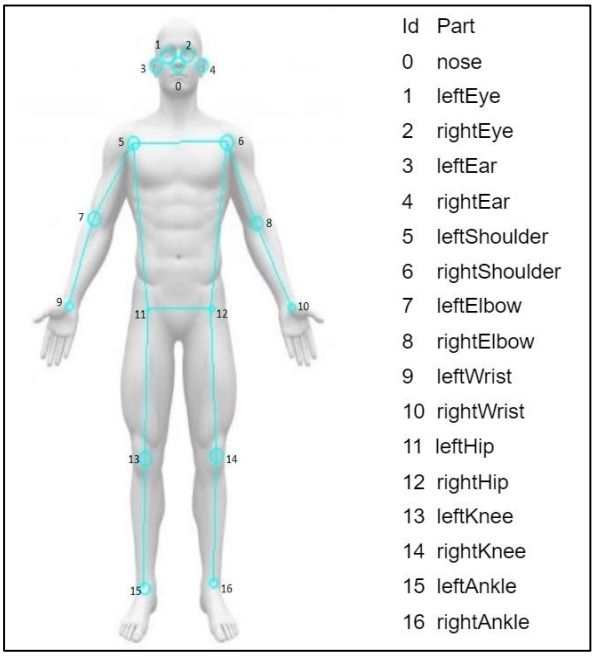

Fig 5. Output Image of Posenet

VI. DNN CLASSIFIER

We have created a DNN to classify different states of an exercise. We receive x,y coordinates of 17 body key points which is then flattened to a single-dimensional array of 34

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 167-173
Published Online May 2020 in IJEAST (http://www.ijeast.com)

coordinates. This array is fed to the DNN classifier which predicts whether the given set of coordinates belongs to state 0 or 1 and gives 0 or 1 as output. The output is then converted to the respective state of the exercise and displayed onto the screen. We have used Binary-Cross Entropy as a loss function and RMSprop as an optimizer. Our DNN consists of 4 layers. In the first layer, there are 128 neurons and then there are 2 hidden layers with 64 and 32 neurons respectively and the output layer has 1 neuron which gives 0,1 as result here relu is used as the activation function. the model was trained on 7660 examples and validation was done on 1916 examples and it was later tested on 2394 examples

VII. RESULT

We present the qualitative and quantitative results of Auto_fit on two different exercises: jumping jack and Lateral shoulder raise. For each exercise, we have prepared a dataset and trained the DNN classifier from scratch.

a) Jumping Jack

Jumping jack is a full-body exercise and is a part of what's called plyometrics, or jump training. Plyometrics is a mixture of aerobic exercise and resistance work. This type of exercise works your heart, lungs, and muscles at the same time. Along with cardiovascular benefits, jumping jacks also offer aids to the lungs. Doing them regularly slowly trains your lungs to expand their capacity, taking in more oxygen and increasing your threshold for physical activity.

Jumping jack can be divided into 2 states:

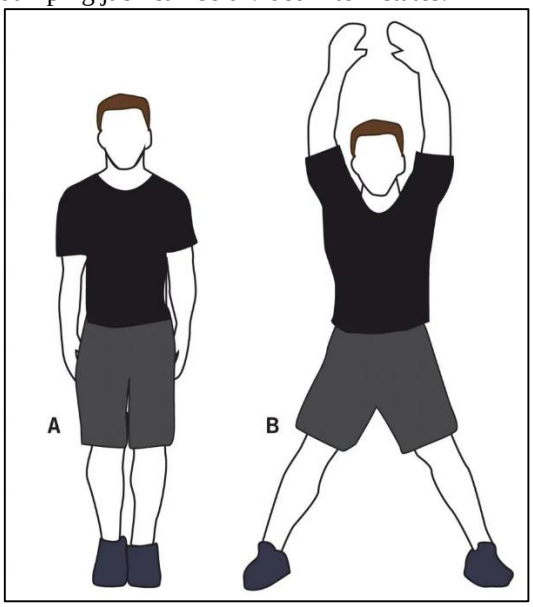

Fig 6. Jumpjack exercise

State A - initial position, in which feet are joined together and hands are down.
State B - from the initial position we do a jump, to move our feet apart and move hands up.
From state B we again take a jump and move hands down and feet close. This cycle keeps on repeating.

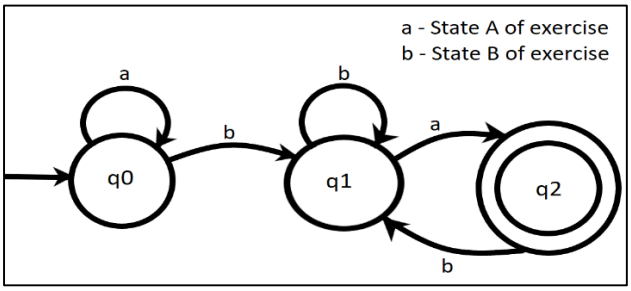

Fig 7. State Diagram for jumpjack

Movement from A to B and again to A is counted as one repetition which means completing and exercising one time. And a collection of a fixed number of repetitions makes a set.

According to MyFitnessPal, jumping jacks can burn around eight calories per minute for a person weighing 120 pounds and up to 16 calories per minute for somebody weighing 250 pounds. For an average person weighing 150lb, he will burn approximately 9 calories per minute of jumping jacks. If he will do 100 repetitions which will take around 2 minutes (depending upon the intensity) then he can burn around 16 calories.
According to WHO[14], Adults aged 18–64 should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity during the week or an equal mixture of moderate- and vigorous-intensity activity. So by only doing jumping jacks for 30 minutes daily, it can fulfill the weekly physical activity requirements.
For creating its dataset we have used videos of trained professionals performing the exercise. We spit the video into 2 different videos containing the state A in one video and state B in the other. Then those videos are passed on to the posenet which generates the coordinates for all 17 key points.

The coordinates from both videos are merged and labeled. This labeled dataset is then split into 3 sets:-

1. Training dataset (7661 examples)
2. Validation dataset (1916 examples)
3. Testing dataset (2395 test examples)

We have trained the model for 100 epochs and gained training accuracy 96% and validation accuracy 95%, training loss 0.12, validation loss 0.28 for loss function binary cross-entropy.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 167-173
Published Online May 2020 in IJEAST (http://www.ijeast.com)

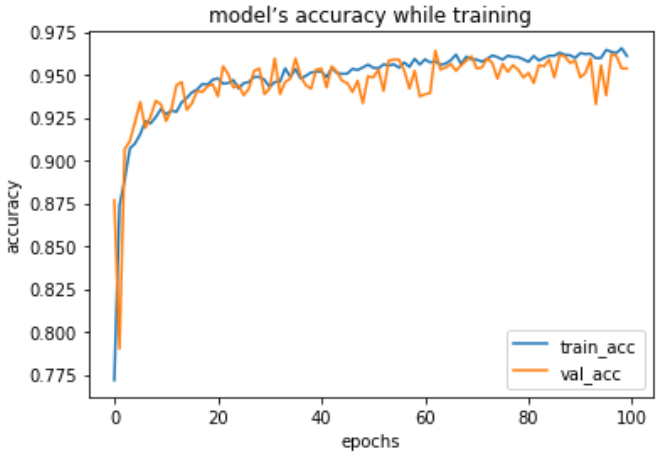

Fig 8. Training accuracy graph

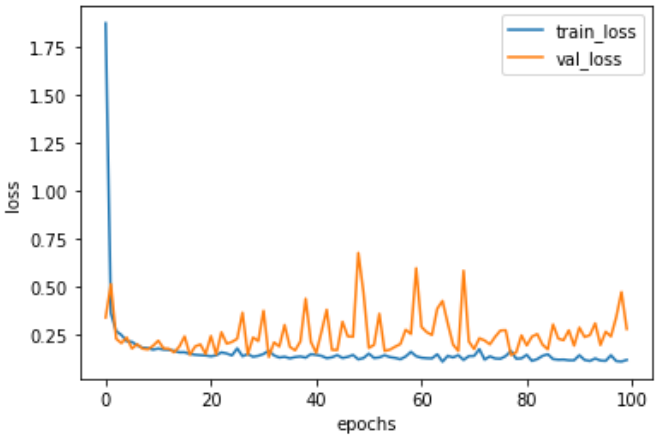

Fig 9. Training loss graph

While running the model on a test dataset with 2395 examples were given we got an accuracy of 95% and loss of 0.25.

b) Shoulder lateral raise

Lateral raise is an isolation exercise that targets deltoid muscles. It is part of a strength workout that focuses on muscle growth. This exercise significantly focuses on the lateral or medial head of the deltoid, creating them seem wider and additional developed. Strength workouts have several advantages like muscle growth, improved bone health, controlled body fat, and minimized Risk of Injury. Due to these benefits, it must be included in a workout regimen.

This exercise can be divided into 2 states :
State A: hands down and feet shoulder-width apart
State B: both hands raised till the shoulder level

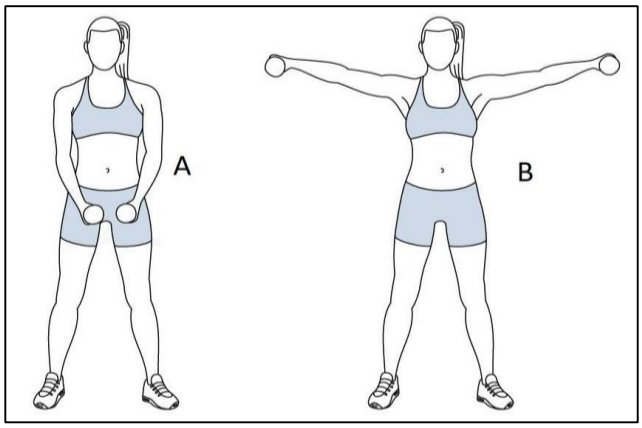

Fig 10. Shoulder lateral raise exercise

The exercise starts with state A, then we gradually start raising arms sideways until reaching the shoulder level. This is state B, then we slowly start lowering the arms and again reach the state A. Movement from A to B and again to A is counted as one repetition which means completing the exercise one time. And a collection of a fixed number of repetitions makes a set.

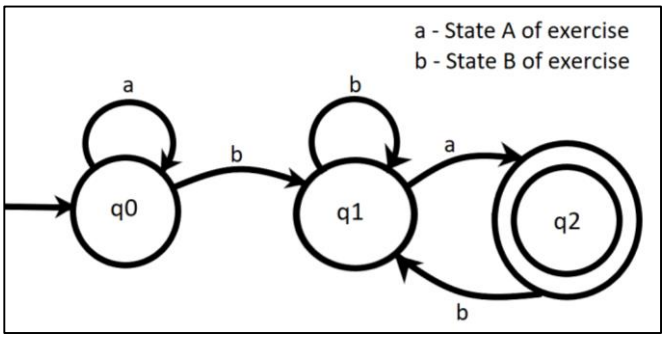

Fig 11. State Diagram Shoulder lateral raise

For creating a dataset for this exercise we have used videos of trained professionals performing the exercise. We have split that video into 2 different videos containing the state A in one video and state B in the other. Then those videos are passed on to the posenet which generates the coordinates for all 17 key points

The coordinates from both videos are merged and labeled. This labeled dataset is then split into 3 sets:-

1. Training dataset (1937 examples)
2. Validation dataset (485 examples)
3. Testing dataset (606 test examples)

We have trained the model for 100 epochs and got Training accuracy 91% and validation accuracy 86%, training loss 0.18, and validation loss 0.15 for binary cross-entropy as a loss function.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 167-173
Published Online May 2020 in IJEAST (http://www.ijeast.com)

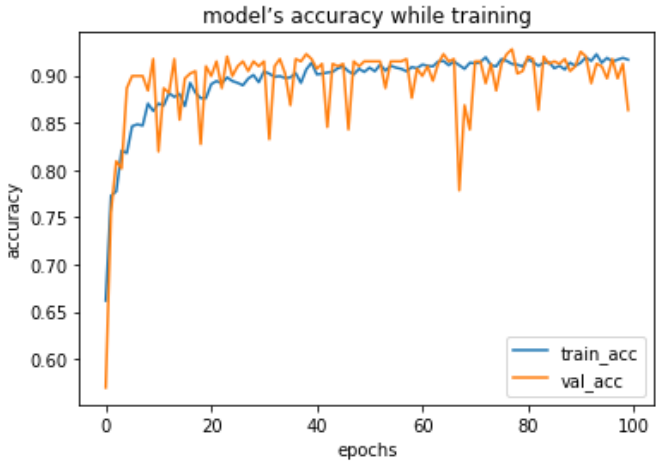

Fig 12. Training accuracy graph

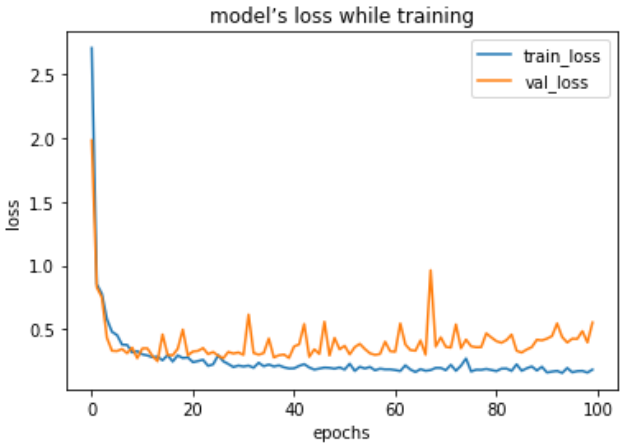

Fig 13. Training loss graph

While testing our model on a test data set we got 86% accuracy and 0.55.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we introduced Auto_fit, an application that uses pose estimation and deep learning to provide effective workout logging and tracking workouts. We have used posenet for pose estimation to evaluate videos of exercises and generate body key points, these are again fed to DNN classifier to identify the state of the exercise. The state information is in turn used for counting the repetitions and sets performed.

We have worked with 2 different exercises, connecting training videos for each, and use both pose estimation and Deep learning to provide repetitions count on a specific exercise, as well as machine learning algorithms to automatically determine the state of exercise in the live camera feed.

Auto_fit does not require very powerful hardware and hence can be easily run on low powered devices like Raspberry pi and Nvidia jetson nano. This enables it's installation in less space and can be produced in a compact form factor.

We have identified several extensions as strong opportunities for future work past this. One way is to build a smartphone-compatible application so it can be easily run on mobile devices and extend its user base so more people can be benefited from this. Another extension is to inculcate more variety of exercises like HIIT and Crossfit workouts. We can also build it's GUI variant so it can be more user friendly and personalized.
According to this review article[1].All-cause mortality is decreased by about 30% to 35% in physically active as compared to inactive and regular physical activity is associated with an increase of life expectancy by 0.4 to 6.9 years. So by using Auto_fit anybody can become physically active and live a longer and healthier life.

IX. REFERENCE

[1] Reimers, Carl & Knapp, G & Reimers, Anne. (2012). Does Physical Activity Increase Life Expectancy? A Review of the Literature. Journal of aging research. 2012. 243958. 10.1155/2012/243958.

[2] O. Patsadu, C. Nukoolkit, and B. Watanapa, (2012)"Human gesture recognition using kinect camera," in Computer Science and Software Engineering (JCSSE), International Joint Conference on. IEEE, 2012, (pp. 28–32).

[3] Pedersoli, Fabrizio & Benini, Sergio & Adami, Nicola & Leonardi, Riccardo. (2014). XKin: an Open Source Framework for Hand Pose and Gesture Recognition Using Kinect. The Visual Computer: International Journal of Computer Graphics. 10.1007/s00371-014-0921-x.

[4] Islam, Muhammad Usama & Mahmud, Hasan & Ashraf, Faisal & Hossain, Iqbal & Hasan, Md. (2017). Yoga posture recognition by detecting human joint points in real time using microsoft kinect. 668-673. 10.1109/R10-HTC.2017.8289047.

[5] Shih-En Wei, Ramakrishna. V, Kanade .T and Sheikh.Y, "Convolutional Pose Machines," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016,doi: 10.1109/CVPR.2016.511.

[6] Cohen .I and Li .H, Inference of human postures by classification of 3d human body shape, in Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on. IEEE, 2003,(pp. 74–81).

[7] Bogo, Federica & Kanazawa, Angjoo & Lassner, Christoph & Gehler, Peter & Romero, Javier & Black, Michael. (2016). Keep It SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image. 9909. 561-578. 10.1007/978-3-319-46454-1_34.

[8] Cao .Z, Simon .T, Wei .S and Sheikh .Y, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, doi: 10.1109/CVPR.2017.143.

[9] Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. In B. Leibe, J. Matas, M. Welling, & N. Sebe (Eds.), *Computer Vision - 14th European Conference, ECCV 2016, Proceedings* (pp. 483-499).

[10] Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. 2017. VNect: real-time 3D human pose estimation with a single RGB camera. ACM Trans. Graph. 36, 4, Article 44 (July 2017), doi:https://doi.org/10.1145/3072959.3073596

[11] Bandhu .A and Roy .S, "Classifying multi-category images using deep learning: A convolutional neural network model," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT),Bangalore,2017,doi:10.1109/RTEICT.2017.8256731.

[12] Jürgen Schmidhuber, (2015) "Deep learning in neural networks: An overview" in neural networks, Elsevier, doi: https://doi.org/10.1016/j.neunet.2014.09.003.

[13] LeCun, Yann & Bengio, Y. & Hinton, Geoffrey. (2015). Deep Learning. Nature. 521. 436-44. 10.1038/nature14539.

[14]https://www.tensorflow.org/lite/models/pose_estimation/overview?hl=ru

[15]https://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html