Exercise Classification using Deep Learning

Abhinav Rawat
IIIT Delhi
abhinav19132@iiitd.ac.in

Ishaan Dayal
IIIT Delhi
ishaan19245@iiitd.ac.in

Dhananjay Singh
IIIT Delhi
dhananjay19161@iiitd.ac.in

April 30, 2022

Abstract

Recent global changes have brought about a huge change in an individual's way of living. People have started realizing the importance of a healthier lifestyle and have become increasingly self aware in terms of diet, exercising, sleeping patterns, etc. However, many individuals lack the requisite resources required for such a lifestyle and it gets much tougher for those who want to start this lifestyle. The aim of this project is to provide a cheap, beneficial way to new gym participants of identifying and learning about new exercises by utilising existing Computer Vision and Deep Learning resources.[1]

1 Introduction

Many sensors can interpret the movements of cardiovascular exercises and give an approximate calculation of calorie burn during the exercise but they do not predict/interpret the movements of a weight training exercise.

1.1 Problem Statement

Our model will take an image of a person performing a weight training exercise as an input and output a description of that exercise, the correct form of that exercise along with the estimated calories burned. To implement this model we are going to use saliency maps for environment detection and segmentation and then digital skeletonization via morphological thinning algorithms as pre-processing on our input image. Our model will be evaluating our final output image using evaluation metrics such as confusion matrix, recall and F1 - score.[2]

1.2 Challenges

- Lack of availability of resources to a certain group of individuals
- High costs of sophisticated wearables that assist in training and exercising
- Lack of prior information to new gym participants resulting in improper exercising routines and high risk of injuries

1.3 Motivation

Youth are always looking for new technologies and cures to help them improve their health and performance. Athletes are increasingly turning to wearable sensors to track their training and recuperation. Wearable technologies currently use sensors and other circuit-based devices which are kind of cumbersome and not that feasible. Moreover, most of the time, newcomers lack the knowledge of identifying and differentiating between different exercises. Using this computer vision based project, any newcomer should be able to identify the type of exercise via a simple image. It could be beneficial for them to know about a certain exercise variation and their descriptions as it can help an individual understand the importance of the aforementioned exercise and the correct way to perform it in order to lessen the chances of injuries[3]

1.4 Contributions

Our contribution with the help of this model will be towards healthcare and general wellness. We plan to help the general public with minimal resources as all that is required for our model to function is a camera picture which anyone can take from their smartphone. This is a model which aims for public good at minimal cost and plans to educate them on weight training exercises by informing them the correct posture and which all muscle groups it targets. This will help people learn about the exercise and prevent injuries as most of the injuries in the gym are caused by incorrect posture. Our model can be expanded into a much more sophisticated tool which can do a deeper analysis with high accuracy, efficiency and provide a detailed output for the user.[4]

2 Summary

Our implementation is divided into three phases - saliency based object segmentation, digital skeleton creation and classification of unknown exercise.

2.1 Deeply Supervised Salient Object Detection

This phase involves detecting the object of interest in the image which using saliency maps. Saliency objection detection was performed using pretrained Deeply supervised model on the dataset which separated the foreground and the background for the creation of binary mask. These images were resized while keeping their aspect ratio same which reduced the resolution and size of the images. The saliency maps created by DL.The binarized saliency maps will segregate the object of interest from the background as shown in Figure 1.

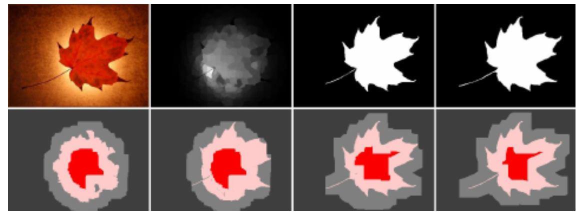

Figure 1. Saliency based segmentation

2.2 Pose Detection

PoseNet is a convolution neural network developed by Google which is a real-time pose detection technique. Posenet is a deep learning TensorFlow model that allows you to estimate human pose by detecting body parts such as elbows, hips, wrists, knees, ankles, and form a skeleton structure of your pose by joining these points. It takes an image as an input and outputs a 16 key point skeleton on the person in the image along with a 3D pose model of that person.[5]

Figure 2. Human Pose Estimation

2.3 Classification of unknown exercise

The skeletons obtained in the first phase will be used training and testing our model over various different exercises and making appropriate classifications

3 Literature Review

Various studies have been conducted to counter this problem with the objective of finding new economically viable options. There have been notable achievements in the field of hardware. Some examples include Invention of sophisticated wearables (as shown in Figure 3) that measure force distributions, hardwares with IMU-based fitness tracking, heart reading, etc. Realizing the importance of smartphones in our daily life, sensor integration with mobiles led to an even more in-depth analysis by taking in parameters such as height, weight, etc which are unique to a user's profile and help in the creation of a fitting database.[6]

However, all these features are advantageous to intermediate to advanced gym participants and are of little significance to those who have recently started. One of the major issues faced by beginners is to be less informed about exercise variations and how a certain variation affects which body parts. This is an even greater issue for the less resourceful people who may not have access to a trainer or appropriate training programs.[7]

Computer vision techniques play a huge role in situations like these as variations can be identified using an image and give a description on it as well. Previous studies have incorporated use of pose estimation techniques for limb identification and utilized algorithms such as Optical Flow Tracking and Dynamic Time Warping(DTW) for motion analysis on video inputs and have shown suitable results. Researchers have experimented from simple classifiers and clustering algorithms such as K-means, Naive Bayes, etc. to neural network based classification such as MLPs and CNNs to identify motions made during exercises.[8]

Figure 3. Fitness tracking wearables

The huge development in the field of detection of salient objects allows using various DL and Non-DL based segmentations and the comparison between their performances. Similarly, performance comparisons between various morphological thinning algorithms allows us to choose the optimal digital skeleton creation algorithm depending upon the need.

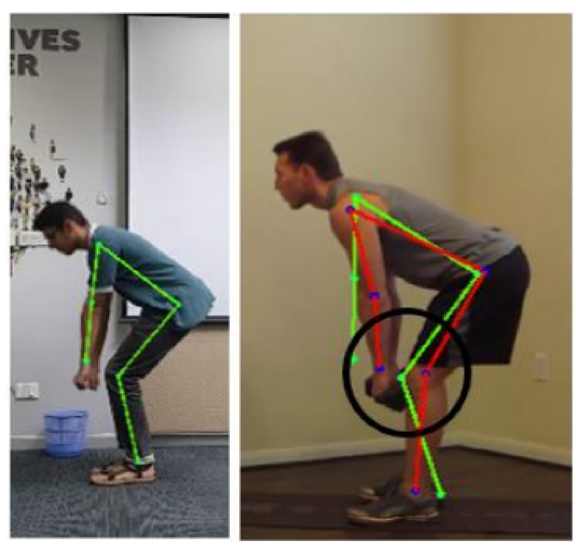

Figure 4. Key Points and Limb detection

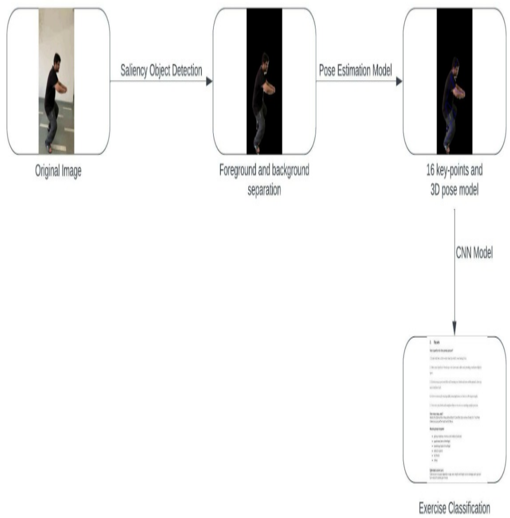

Figure 3. Model Pipeline

4 Methodology

4.1 Dataset

The image dataset was video based on 3-5 exercises and was made under the appropriate environment. The video was broken down into 4358 frames which were converted into numerous images and divided on the basis of the type of exercises and split into approximately 25:15:3 into train, validation and test sets respectively. The images were of resolution 1280 x 720 pi which is later reduced to 160 x 90 pi

4.2 Salient Object Detection

Saliency detection methods based on multi scale featured maps and introduction of short connections to the skip-layer structures within the Holistically-Nested Edge Detector (HED) architecture have been used for background subtraction for removing noise from our dataset to improve the accuracy of the model. The images that we get as the output, .i.e without the background are brought down to a lower resolution of 160 x 90 pi while maintaining the aspect ratio of 16:9 at the same time.

4.3 Pose Estimation

The pipeline uses a state-of-the-art 2D pose estimation sub-network with a 3D depth regression sub-network for pose estimation. This model took the resized images as its input and output was images with 16 key points in the Cartesian plane and a 3D pose of that image.

4.4 Deep Convolution Network Classification

A deep learning model was created using CNN approach. This model took foreground images as its input and performed a series of convolution layers to give a one-dimensional feature vector. The key points information was first flattened into a single dimension vector and then appended on to the one-dimensional feature vector which was finally sent through a linear layer and a Softmax activation to give probabilities for each class which helped classify the exercise.

5 Results

On performing the experiment, we observed a training accuracy of around 98.2 and a validation accuracy of around 97.87. The test set gave an accuracy of 97.48

5.1 Loss Plots

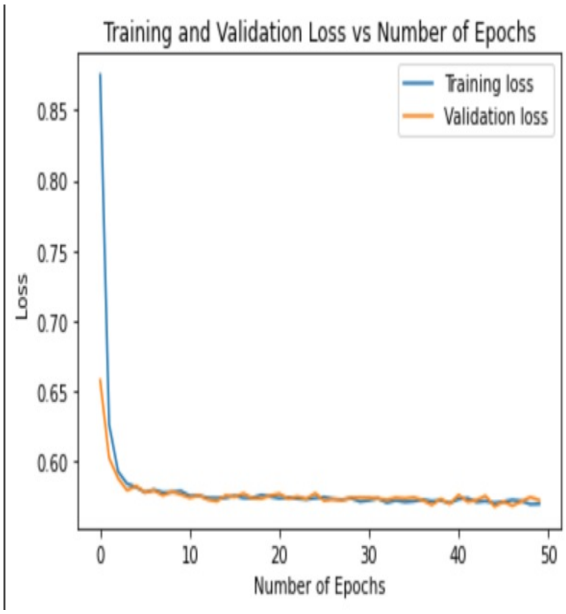

Figure 5 Training and Validation Loss

5.2 Accuracy Plots

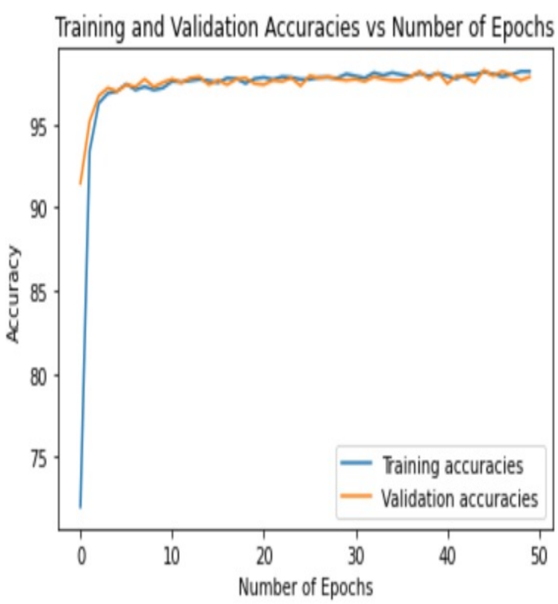

Figure 6 Training and Validation Accuracy

5.3 Classification Report

	precision	recall	f1-score	support
0	1.00	0.92	0.96	83
1	0.95	0.98	0.96	123
2	0.99	1.00	0.99	151
accuracy			0.97	357
macro avg	0.98	0.97	0.97	357
weighted avg	0.98	0.97	0.97	357

Figure 7 Classification Report

6 Conclusion

From the experiments and training of our model, we observe that enriching the image data with key points data helps a lot in improving the accuracies, precision, recall and F1 score as observed in Figure 7. Hence, it can be concluded that the model was successfully implemented and gave good classification results on our data.

References

[1] Wei Chen, Lichun Sui, Zhengchao Xu, and Yu Lang. Improved zhang-suen thinning algorithm in binary line drawing applications. In *2012 International Conference on Systems and Informatics (ICSAI2012)*, pages 1947–1950. IEEE, 2012.

[2] Zicheng Guo and Richard W Hall. Parallel thinning with two-subiteration algorithms. *Communications of the ACM*, 32(3):359–373, 1989.

[3] Ashish Kumar Gupta, Ayan Seal, Mukesh Prasad, and Pritee Khanna. Salient object detection tech-

niques in computer vision—a survey. *Entropy*, 22(10):1174, 2020.

[4] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip HS Torr. Deeply supervised salient object detection with short connections. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3203–3212, 2017.

[5] Usman Ali Khan, Iftikhar Ahmed Khan, Ahmad Din, Waqas Jadoon, Rab Nawaz Jadoon, Muhammad Amir Khan, Fiaz Gul Khan, and Abdul Nasir Khan. Towards a complete set of gym exercises detection using smartphone sensors. *Scientific Programming*, 2020, 2020.

[6] Ming Li, Jiangying Qin, Deren Li, Ruizhi Chen, Xuan Liao, and Bingxuan Guo. Vnlstm-posenet: A novel deep convnet for real-time 6-dof camera relocalization in urban streets. *Geo-spatial Information Science*, 24(3):422–437, 2021.

[7] Amit Nagarkoti, Revant Teotia, Amith K Mahale, and Pankaj K Das. 1. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1440–1443. IEEE, 2019.

[8] Aleksandr Ometov, Viktoriia Shubina, Lucie Klus, Justyna Skibińska, Salwa Saafi, Pavel Pascacio, Laura Flueratoru, Darwin Quezada Gaibor, Nadezhda Chukhno, Olga Chukhno, et al. A survey on wearable technology: History, state-of-the-art and current challenges. *Computer Networks*, 193:108074, 2021.

[9] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. Deep 3d human pose estimation: A review. *Computer Vision and Image Understanding*, 210:103225, 2021.

[10] Bo Zhou, Mathias Sundholm, Jingyuan Cheng, Heber Cruz, and Paul Lukowicz. Measuring muscle activities during gym exercises with textile pressure mapping sensors. *Pervasive and Mobile Computing*, 38:331–345, 2017.

[11] Xingyi Zhou, Qixing Huang, Xiao Sun, Xiangyang Xue, and Yichen Wei. Towards 3d human pose estimation in the wild: a weakly-supervised approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 398–407, 2017.