# Project Proposal

## Project Title: Posture Correction Using Computer Vision

## Group Name: Exposed-Shots


## Group Members: Ayushi, Gurneet Chhabra


## Problem Statement:

Fitness and Technology are two of the fastest-growing industries in the world. Fitness reeled in an estimated $94 billion last year at an annual growth rate of 6.1%. Technology permeates the world, with the AI industry reigning as one of the fastest growing and the intersection of these two industries has created a whole new world of digital fitness. This project is a Personal Workout AI Assistant based on employing AI and computer vision to build a personal fitness trainer. The assistant will help the user in performing the exercises with maximum efficiency with other added features like Reps, time, calories, etc.

Recent progresses in fields of pose estimation, action recognition and motion prediction allow us to analyze movements in details and thus identify **potential mistakes** done while exercising. In this work, we prepare a dataset containing videos, 2D and 3D poses of correct and **incorrect executions** of different movements that are **Squats, lunges, planks and pick- ups** and labels identifying the mistake in each practice of that exercise. This dataset is used to demonstrate our motion correction model, designed using a **Graph Convolutional Network architecture** and trained with a differentiable dynamic time warping loss. As a result, we are able to correct movement mistakes in 3D pose sequences and output the corrected motion. This model is integrated in a pipeline containing a state-of-the-art 3D human pose estimator to go from raw video images to a sequence of corrected 3D poses. Evaluation of this model is done using an action recognition model trained on the same dataset to recognize whether the sequence is correct or has a particular type of mistake.

## Related Work:

Our motion correction task lies at the interface of computer sciences and human motion analysis. This is also the case of the two other tasks that are **action recognition (AR)** and **motion prediction (MP).** As in our case, they take as input a sequence pose (or images) and aim to analyze it, whether to assign it a label in the case of AR or to forecast future body poses in the case of MP. Most of the time, AR is done to discriminate between a limited number of actions that are often easily distinguishable one from another for a human eye. For example, UCF101 [1] and HMDB-51[2] datasets actions can be sports (horse riding, skiing, rowing, ...), everyday activities (brush hair, type, cooking, ...), simple movements (clap, throw, turn, ...) and many others.

One way to improve AR performance has been recently investigated by Li et al.[3]. In this work they show that fusing multiple modalities, and in particular pose information, can improve classification accuracy and even reach state-of-the-art results. AR does not necessarily work with pose information, MP on the other hand needs a structured input in order to output something of the same shape. In their

two papers, Mao et al. [4, 5] propose a Graph Convolution Networks (GCN) structure encoding the spatial representation of the motion. [5] improves upon the state-of-the-art scores of [4] by adding an attention mechanism capturing motion redundancies. Due to their successful handling of 3D human motion sequences, we have based our motion correction architecture as a GCN as proposed in these works.

## Dataset:

In the frame of this project, we have formed a new dataset containing both correct and incorrect versions of a physical exercise, both executed by the same subject. the data is acquired in a room equipped with 4 GOPROs approximately positioned on a circle around the subject performing the actions standing in the middle. The cameras are oriented so that the subject appears in the middle of the image. The sampling rate is 30 frames/second and the images have a size of 1920x1080 pixels.

The list of actions to be performed by the subject, as well as the instructions and the number of repetitions, are summarized in table 1. This process is repeated for 4 different subjects but all the data is acquired the same day under the same conditions. SQUATs, lunges and planks are strength exercises.

## Timeline:

The data collection will take about 2 weeks. Here we are building our custom dataset. Implementation of **action recognition (AR)** and **motion prediction (MP)** model will take 2 more weeks. Testing and Validation of the model's performance will take one week.

## Reference:

[1] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. (November), 2012.

[2] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A large video database for human motion recognition. Proceedings of the IEEE International Conference on Computer Vision, pages 2556–2563, 2011

[3] Yinxiao Li, Zhichao Lu, Xuehan Xiong, and Jonathan Huang. Perf-net: Pose empowered rgb-flow net. arXiv, 2020.

[4] Wei Mao, Miaomiao Liu, Mathieu Salzmann, and Hongdong Li. Learning trajectory dependencies for human motion prediction. Proceedings of the IEEE International Conference on Computer Vision, 2019-Octob:9488–9496, 2019.

[5] Wei Mao, Miaomiao Liu, and Mathieu Salzmann. History repeats itself: Human motion prediction via motion attention. arXiv, 2020