[1]Morarjee Kolla

[2]Phani Varma Gadiraju

[3]Dhruvraj Tondanoorthy

Form Check: Exercise Posture Correction Application

Journal of Electrical Systems

Abstract: - Weightlifting is a popular sport and a hobby of millions; however, the risk of injury is high, primarily when performed with incorrect form/posture due to voluntary or involuntary reasons. It results in a significant loss of strength and hinders progress. The gradual strain on tendons and ligaments causes wear and tear. This paper aims to develop a deep learning model that detects incorrect posture and provides insight into remedial steps that need to be taken to correct the same, thereby providing increased safety and efficiency. Our Form Check model detects a user's form, vector geometry of posture is evaluated and corrected. A graphical representation of the exercise is provided as feedback to the user. Object detection, pose estimation, and action recognition algorithms infer the human skeleton and recognize the exercise. The inconsistency in the exercise performed is determined and mapped to specific imperfections in the stance of users during the exercise through Graph Convolution Networks (GCN). Form Check is designed for exercises with a high potential risk of injuries like squats, lunges, and planks.

Keywords: Action Recognition, Form Check, Graph Convolution Networks, Pose Estimation.

I. INTRODUCTION

Form Check is an exercise posture correction model that can detect, recognize the activity and correct the exercise a user is performing. While we implemented Form Check for weightlifting exercises here, it can find use in gyms, sports, and yoga. Any user, from a trainee to a veteran, can use Form Check to make fine adjustments to their exercise posture. This helps reduce the risk of injury and increase efficiency.

Pose Estimation and Human Activity Recognition are two essential segments in Form Check. Human pose estimation localizes body key points to accurately recognize the postures of individuals given an image [1]. Activity recognition mimics the human ability to recognize another person's activities [2].

Form Check uses You Only Live Once (YOLO) [3] as a person tracker, VIBE [4] for pose estimation, and Graphical Convolution Network (GCN) [5] based models for exercise classification and correction. A graphical representation of performed and corrected exercises is displayed.

Weightlifting is the body's movement produced by muscular tissues contracting, resulting in the motion of limbs or torso. It generally involves using barbells, dumbbells, or just body weight. Weightlifting aims to improve strength, build muscle mass, reduce weight, and improve energy levels and persistence. The exercises involved can be beneficial to the health of individuals. However, it is a potential risk for injury and would be inefficient in attaining the required results when performed incorrectly. Newcomers to this field are especially prone to this risk. However, bad posture is something all levels of gym-goers occasionally fall prey to.

It is difficult to perfect the complicated series of actions involved in weightlifting exercises. Posture problems can manifest into muscular and joint conditions, putting a trainee at risk in areas such as the shoulder, neck, spine, and knees. Proper posture can reduce the strain on the human frame by preserving the integrity of tissues and the skeleton. A balanced musculoskeletal motion may be essential to prevent further damage and promote healing in weightlifting. Therefore, learning the correct posture is essential to any exercise, especially for people lifting heavy weights, which is quite common among gym-goers.

[1] [1]Associate Professor, Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

morarjeek_cse@cbit.ac.in

[2,3]B. E. Student, Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

ugs18034_cse.phani@cbit.org.in, ugs18028_cse.dhruvraj@cbit.org.in

II. LITERATURE SURVEY

After going through research papers, publications, and other academic resources, we have compiled a list of technologies and described technologies relevant to our work. These resources are referred to and cited as it has contributed to designing and implementing our proposed system.

Lei Yang et al. [6] proposed a paper titled "Human Exercise Posture Analysis based on Pose Estimation," intending to provide a convenient way to learn and correct the techniques during exercise; the authors propose software to analyze the posture with just a camera. Users' movements are tracked, and exercise analyses are performed to guide them.

A user can use/upload any video. The pose estimation method that follows a bottom-up approach - OpenPose [7] is used for Pose Estimation, as shown in Figure 6. To improve the Pose Estimation accuracy of Open Pose, the authors suggested using regularization techniques on the dataset used for training OpenPose.

During Pose Estimation Phase, the authors discovered that in unique scenes involving non-upright people, occlusion, and human truncation, OpenPose produced wrong results. Data augmentation and multi-directional input through the rotation of images are used to solve this problem.

For the analysis of squats, the authors considered the angles formed between various key points as a metric for determining the correct form. The mathematical expression is of the form: $\angle$ (x1, y1), (x2, y2) $\in 90° \pm (\theta)$, where x1, y1, x2, y2 are the coordinates of hips, knees, and feet, and θ is the error range allowed.

In the case of push-ups, the height of the hips is the criterion. The correct position of the hips should be the average height of the shoulders and ankles, which is mathematically defined where the coordinates of the hip, shoulder, and feet are known. The output involves feedback to the users consisting of the analysis results stating whether they performed the exercise correctly.

Mahendran N et. Al [8] proposed a paper titled "Deep Learning for Fitness" in which they use a novel pose estimation technique developed by TensorFlow, Pose Net, to develop a FitnessTutor. This Real-time exercise posture comparison application guides the users toward the correct form. This application can work on any exercise, provided there is a reference image to compare the user's posture.

The author proposes a method of correcting the posture with and without a coach. The deep learning PostNet [9] model is used to help correct the posture. The coach-provided training image can be compared with the current posture. If the key point positions are wrong, the PostNet model could identify with the comparison with the coach's posture, as shown in Figure 7, and tell the user what they need to correct. A mathematical approach is used for the comparison of images. The referenced skeleton is matched with the user skeleton generated in real-time for suggesting the corrections.

Steven Chen et. Al [10] proposed a paper titled "Pose Trainer: Correcting Exercise Posture using Pose Estimation" in which they proposed an application titled 'Pose Trainer'. This application can detect the user's exercise pose and provide helpful feedback on the user's form to aid users in performing exercises in the correct posture. The latest pose estimation and machine learning algorithms were used to achieve this.

Initially, a pose estimation model uses an RGB image or depth map as visual input data infers the human pose, and determines a person's joints. Figure 8 shows a list of skeletal key points. For this inference, a state-of-the-art deep neural network-based pose estimation model at that time, OpenPose, was used within Pose Trainer. Open Pose uses the first ten layers of VGG-19 [11] for feature extraction and two branches for finding Part Affinity Fields (PAFs) and Confidence maps.

The quality of a user's pose is detected and predicted for the given input exercise pose in the second part of the Pose Trainer. A heuristic machine learning model is further responsible for determining the ground truth in proper form using the instructions and poses of trained professionals, personal trainers, and other qualified professionals.

The second stage involves the evaluation of exercise posture given normalized key points. Due to the arbitrary length of recorded videos, this results in a different key point vector length for each example. Different vector lengths of features present a challenge for many machine learning models. The approach to this task is through dynamic time warping (DTW) with the nearest neighbor classifier. DTW tries to dynamically identify the key point corresponding to the given point in the first sequence in the second sequence. Feedback is given to the users based on inconsistencies in the posture as a text message. The summary of all the existing solutions discussed above is presented in Table 1.

Table 1. Summary of existing solutions

S.No	Title	Methodology	Advantages	Disadvantages	Remarks
1	Pose Trainer: Correcting Exercise Posture using Pose Estimation	• Uses the first few layers of VGG-19 for feature extraction • Dataset was augmented and given as a multi-directional input • The angle between joints was the metric of evaluation	• Fast • No Additional Post-Processing required • Runs in real-time required	• Errors in the Pose Estimation phase due to occlusion will disruptive • It is not suited for all audiences as angles are not a good metric for evaluation	• The use of Machine Learning Algorithms in correcting phase will significantly improve the accuracy
2	Deep Learning for Fitness	• Uses PostNet for pose estimation • Compares the pose of a user to an ideal pose given as input	• Fastest approach to pose correction • Lightweight and easy to deploy	• Requires some level of human intervention/input	• Verbal and Visual feedback, along with the availability of a dataset for exercises, will make the idea feasible
3	Human Exercise Posture Analysis based on Pose Estimation	• Uses VGG-19 for feature extraction and two branches of CNN for PAFs and Confidence Maps • Uses heuristic-based machine learning model (DTW) to determine ground truth	• Precision is high for most exercises. • The dataset used for training creates a well-balanced model	• Slow, cannot be implemented in real-time • The lack of an activity recognition phase restricts the model to just one exercise at a time.	• The use of a faster pose estimation model like PostNet and activity recognizer will allow it to be deployed in real time.

Based on the comparisons of all the solutions present, we find that these methods rely on state-of-the-art Pose Estimation models present at the time of their creation and their application primarily using mathematical functions. These functions fail to consider the inconsistencies present in input data accurately. This restricts the videos that can be used as input.

Another vital area lacking in modern posture correction applications is the presence of an action recognition model. This model can automatically detect the type of exercise being performed, reducing the need to input the same manually and allowing for a faster and more streamlined experience when using a posture correction application. Finally, visual feedback is missing, and the feedback is primarily through text messages.

III. PROPOSED METHODOLOGY: FORM CHECK

The design and workflow of Form Check are described in this section. There are multiple modules involved in the implementation of Form Check. The information about the proposed system, its design, the fundamental representation of the process flow, and the functionality of each module are described in this section. We also show the algorithms used to make the system that aided the whole development of our design.

1.1 Functional process

The flow of the block diagram is developed in a way to address each module in sequence. Initially, the input Video is given by the user (uploaded in the data folder in Google Drive). Then, the video is processed by the multi-person tracker, which sends the (.pickle) file to the Pose Estimator algorithm. Now, this algorithm classifies the activity inside the video, searches the dataset for the exercise, and provides the result to the Posture Corrector. Later, the Posture Corrector algorithm corrects the wrong or incorrect posture analyzed by the activity recognizer and gives a vector geometry plot on a 3D plane as shown in Figure 1.

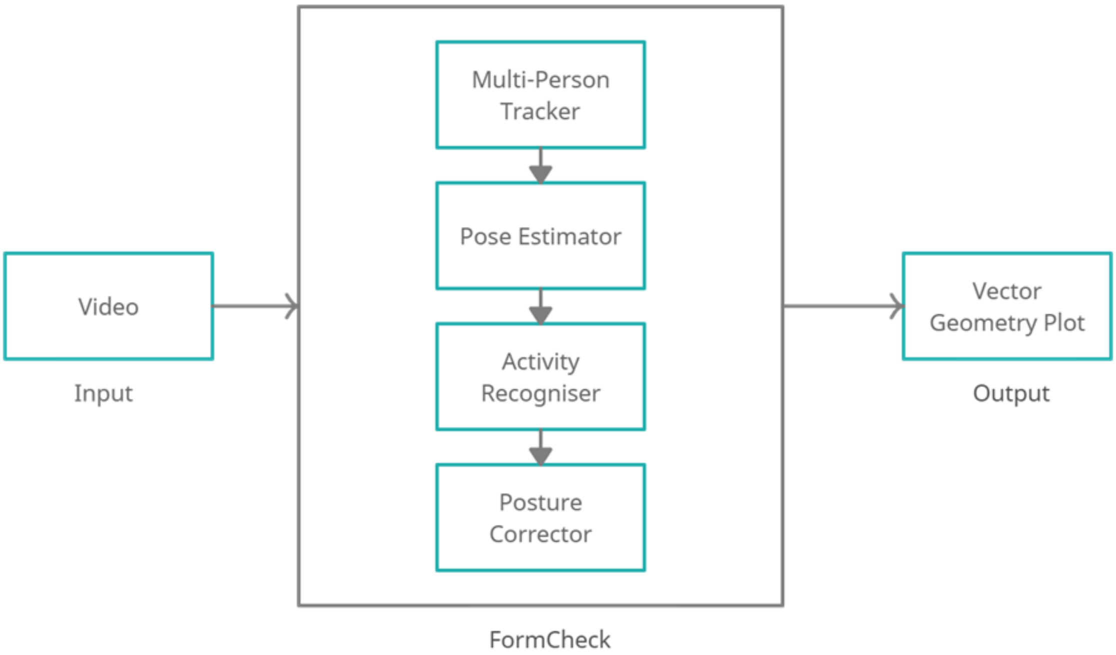

Figure 1: Functional Process of Form Check

First, we use the multi-person tracker library, which contains You Only Live Once (YOLO) as a part that tracks a person through object detection. Second, we use Video Inference for Human Body Pose and Shape Estimation (VIBE) to convert the output from YOLO to a 3D representation of the human pose. The final and most crucial stage of Form Check involves using a classification and correction model to determine the type of exercise performed and correct it accordingly. We plot the results using matplotlib. Finally, we represent the output posture in two colors, red being the input posture (either wrong or incorrect) and green being the correct and ideal posture.

1.2 Classification Model

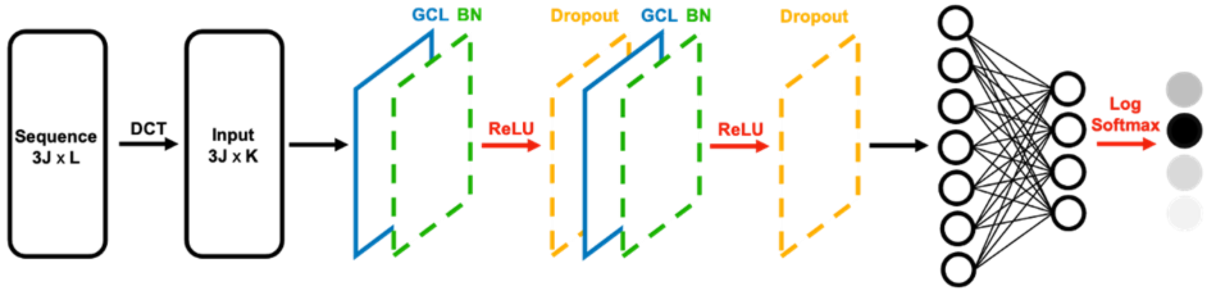

Figure 2: Network Diagram of Classification Model - Form Check

The sequence starts by taking the input video and uses DCT [12] to provide the input to the GC Block. The function of a discrete cosine transform (DCT) is to express a finite sequence of data points as a sum of cosine functions oscillating at various frequencies. A Graph Convolutional Network (GCN) is a semi-supervised deep learning technique used on graph-structured convolutional data. It is an efficient variant of CNN that can operate directly on graphs. Graph Convolutional layer assumes a constant input graph composition passed as a layer altercation. As a result, the input order of graph nodes is fixed for the model and should match the nodes' order in inputs. We define a GC Block and pass our input through it. The following process is shown in Figure 2 as an architectural diagram.

One Graph Convolutional Block (GC Block) consists of the following segments,

- GCL (Graph Convolutional Layer)
- BN (Batch Normalization) [13] – faster method of normalization
- ReLU (Rectified Linear Unit) – activation function
- Dropout (or Dilution) [14]

LogSoftmax is an activation function used after the fully connected layer to classify the activity performed by the person in the input video.

1.3 Correction Model

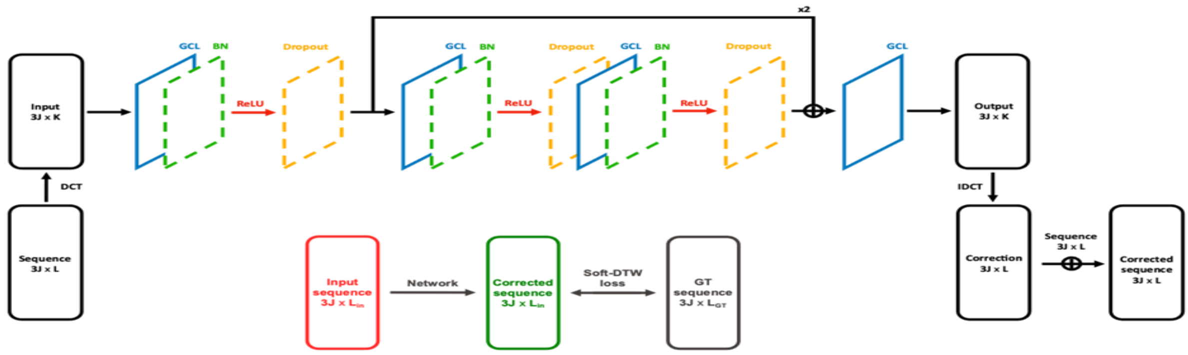

Figure 3: Network Diagram of Correction Model - Form Check

In this model, we use 3 Graph Convolutional Blocks, as shown in Figure 3. The input sequence is sent to the first GC Block, and then it is split into two different branches. The first branch passes the input through two additional GC Blocks and adds it to the previous workflow. Then this is passed through another graph convolutional layer to produce the output where Inverse Discrete Cosine Transform (IDCT) is performed to get the output time series data containing the correction. This correction is added to the input sequence to produce the correct sequence. Finally, the corrected sequence uses Soft Dynamic Time Warping [15] (SoftDTW) loss to calculate the Euclidean distance between each key point mapped from the estimated correct sequence and the input sequence. Now this corrected output sequence is shown in green with the input sequence shown in red. Body Posture Detection and Motion were observed with AI approaches in our system [16] [17].

IV. IMPLEMENTATION OF FORM CHECK

Form Check is implemented in Jupyter Notebook and executed in Google Colab. This ensured a similar execution environment with all the correct dependencies. The data and required pre-requisite models like YOLO and VIBE can be downloaded onto any drive and mounted into the Colab notebook.

Google Drive is accessed, and the contents required are given to Google Collaboratory (Colab). We then run all the prerequisites and requirements to check the version compatibility and libraries. After everything is up and running without errors, we run all the modules in sequence.

We import the deep learning models and functions required to define our network from PyTorch and TensorFlow. The input videos are stored in this drive to be assessed.

The implementation takes a given video input, breaks it into individual frames, and passes them through each model. The person tracking and object detection data is stored in pickle format. This data object is used as input for the classification and correction model. As the data involves a graphical representation of human key points, GCN is used for classification and correction. The required results are obtained here.

Finally, the graphical output is plotted. The corrected posture is observed in green, and the wrong/incorrect (input) posture is observed in red. The user can further check any other exercise video by uploading the same to the associated drive.

The dataset used contained 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. We used HV3D's video dataset to train the model in the fields of 3D Human Postures and Exercise activities. The input videos are given through the Google Drive data directory.

V. RESULTS AND DISCUSSION

Form Check was run to correct a few training samples of different exercises consisting of correct and wrong techniques, as mentioned in Table 2. The exercises include squats, lunges, and planks. Some possible ways to perform an exercise correctly include keeping knees wider than shoulder length, arching back, and not going down low enough.

Table 2. Exercises and repetitions performed

Action	Instruction	Repetitions
SQUATs	Correct	10
	Feet too wide	5
	Knees inward	5
	Not low enough	5
	Front bent	5
Lunges	Correct	10
	Not low enough	10
	Knee passes toe	10
Planks	Correct	10
	Arched back	10
	Hunch back	10

The output given by form check for a squat is shown in Figure 4. It represents how the posture displays the vector geometry of the human skeleton and gives us visual feedback. The distance between green and red lines represents the difference between the ideal and wrong positions. In contrast, the left half shows low deviations in the exercise performed.

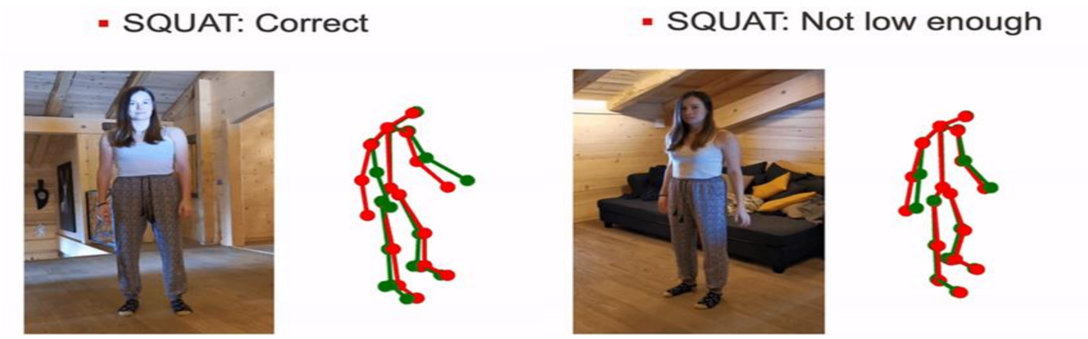

Figure 4: Correct and Incorrect Squat Posture

Figure 5 shows how the skeleton Figure consisting of a human pose looks like for various proper and improper techniques. When an improper technique is used, the difference between key points increases based on which we can determine what went wrong.

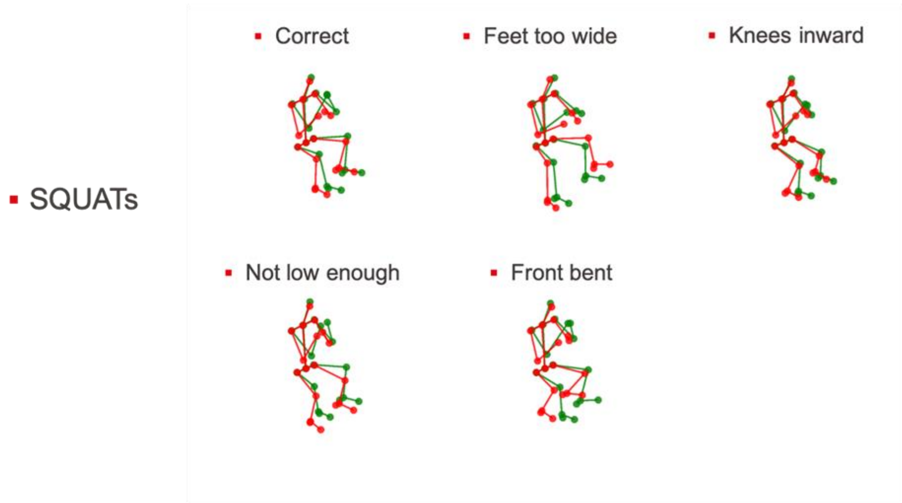

Figure 5: Different Correct and Incorrect Squat Postures

Figure 6 shows a proper and improper method to perform lunges. The left half represents an adequately executed technique, while the right half is improper. Lunges is an exercise where a person extends one of his legs forward and performs a squat such that the bent knees form a square.

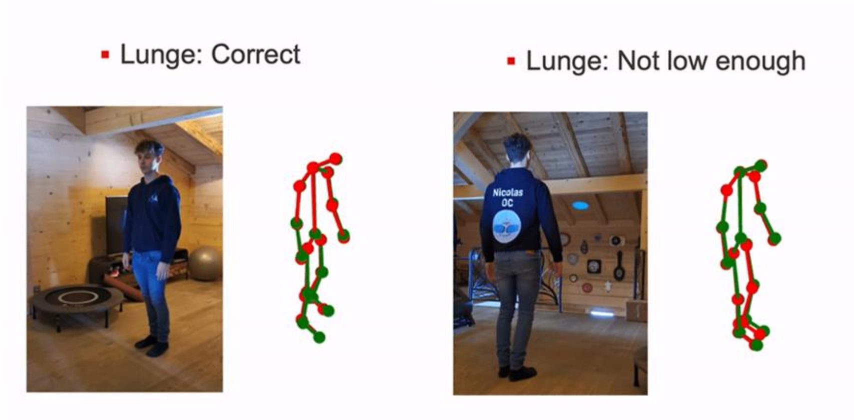

Figure 6: Correct and Incorrect Lunges Posture

Different techniques and positions through which proper lunges and planks could be improperly performed are shown in Figure 7.

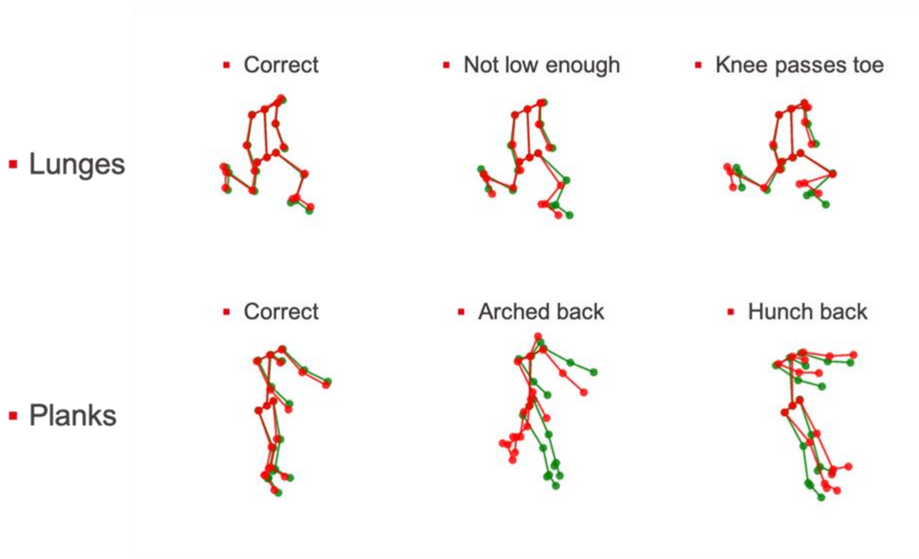

Figure 7: Different Correct and Incorrect Lunges and Plank Posture

We used a dataset with a combination of both proper and improper posture. Although most exercises are predicted with a high degree of accuracy, some movements like the correct squat, forward bent squat, and not low enough lunges are not perfectly accurate. This is likely because the difference is too small to be noticeable and could be considered a borderline between right and wrong. Increasing the margin of error and considering other human body features like height and stature could mitigate this.

VI. CONCLUSION

Pose estimation and activity recognition have become highly researched in computer vision over the last decade. With the ever-increasing performance capacity of modern-day computers, these tasks have only been getting faster and more accurate. In this project, we proposed 3D Human Posture Correction, which guides and aids people in how to correct their posture during exercise. We proposed a deep learning model, Form Check consisting of the GCN architecture. We made the classifier and corrector algorithms using these three methodologies: Multi-Person Tracker, YOLO, and VIBE. Form Check application provides feedback to the user based on their posture during the exercise and informs them of corrective measures. When a video of the user performing an exercise is given to the model, it localizes the joints through key point mapping, classifies the exercise, and then corrects it. The Heuristic Machine Learning algorithm of Soft Dynamic Time Warping is used to determine whether the task is being performed correctly, allowing the user to change accordingly. Our project successfully produces a graphical visualization of all the proper and improper postures. In the future, the primary space can be developed using a quicker Key point Detection Model, which reduces inference time for pose estimation through faster mapping of key points. Real-time video can also be integrated into our system, aiding the users in pose estimation by using their cameras for posture correction. The evaluation can be faster with a more precise model with access to more APIs and the latest resources for object detection. Our model performance can be improved using 3D mapping APIs instead of 2D.

REFERENCES

[1] Munea, Tewodros Legesse, Yalew Zelalem Jembre, Halefom Tekle Weldegebriel, Longbiao Chen, Chenxi Huang, and Chenhui Yang. "The progress of human pose estimation: a survey and taxonomy of models applied in 2D human pose estimation." *IEEE Access* 8 (2020): 133330-133348.

[2] Vrigkas, Michalis, Christophoros Nikou, and Ioannis A. Kakadiaris. "A review of human activity recognition methods." *Frontiers in Robotics and AI* 2 (2015): 28.

[3] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788. 2016.

[4] Kocabas, Muhammed, Nikos Athanasiou, and Michael J. Black. "Vibe: Video inference for human body pose and shape estimation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5253-5263. 2020.

[5] Zhang, Si, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. "Graph convolutional networks: a comprehensive review." *Computational Social Networks* 6, no. 1 (2019): 1-23.

[6] Yang, Lei, Yingxiang Li, Degui Zeng, and Dong Wang. "Human Exercise Posture Analysis based on Pose Estimation." In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, pp. 1715-1719. IEEE, 2021.

[7] Cao, Zhe, Tomas Simon, Shih-En Wei, and Yaser Sheikh. "Realtime multi-person 2d pose estimation using part affinity fields." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291-7299. 2017.

[8] N, Mahendran. “Deep Learning for Fitness.” arXiv preprint arXiv:2109.01376 (2021).

[9] Kendall, Alex, Matthew Grimes, and Roberto Cipolla. "Postnet: A convolutional network for real-time 6-DOF camera relocalization." In *Proceedings of the IEEE international conference on computer vision*, pp. 2938-2946. 2015.

[10] Chen, Steven, and Richard R. Yang. "Pose Trainer: correcting exercise posture using pose estimation." *arXiv preprint arXiv:2006.11718* (2020).

[11] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[12] Ahmed, Nasir, T_ Natarajan, and Kamisetty R. Rao. "Discrete cosine transform." *IEEE transactions on Computers* 100, no. 1 (1974): 90-93.

[13] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In *International conference on machine learning*, pp. 448-456. PMLR, 2015.

[14] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15, no. 1 (2014): 1929-1958.

[15] Cuturi, Marco, and Mathieu Blondel. "Soft-dtw: a differentiable loss function for time-series." In *International conference on machine learning*, pp. 894-903. PMLR, 2017.

[16] Patil, Anuj, Darshan Rao, Kaustubh Utturwar, Tejas Shelke, and Ekta Sarda. "Body Posture Detection and Motion Tracking using AI for Medical Exercises and Recommendation System." In *ITM Web of Conferences*, vol. 44, p. 03043. EDP Sciences, 2022.

[17] Angosto, Salvador, Jerónimo García-Fernández, and Moisés Grimaldi-Puyana. "A systematic review of intention to use fitness apps (2020–2023)." *Humanities and Social Sciences Communications* 10, no. 1 ,2023.