International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Virtual Coach: Monitoring Exercises and Aerobic Dance Generation

Kruti Pandya[1], Aditya Singh[1], Saurabh Pande[1], Dr. Shubhangi Vaikole[2]

[1]Student, Dept. of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India
[2]Professor, Dept. of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

---

Abstract - *Virtual Coach was developed to play the role of a personal fitness trainer. The main aim was to prevent the trainee from injury caused by incorrect exercise form. This was achieved using Pose Estimation. The sequence of poses captured while performing a particular exercise was mapped to the correctness of the form. This information about correctness was then used to provide real-time captions, audio and visual feedback to the user. This sequential model was trained with data comprising weightlifting and muscle building exercises. Since yoga involves holding the body in a particular position, the assessment for yoga was done using a geometric model which compared the user pose to a reference pose. The second aim of Virtual Coach was to make cardio fun by generating aerobic dance moves dynamically on beats of the user's choice. The LSTM model mapped music features to dance pose key points extracted by Pose Estimation. These 2D dancing skeletons were then rendered on the screen for the user to follow along. The dance moves generated were natural and harmonious. Results demonstrated that Virtual Coach is reliable and capable of assisting the trainee in maintaining the correct exercise form during workout.*

Key Words: **Pose Estimation, exercise form, real-time, sequential model, yoga, weightlifting, dance.**

1.INTRODUCTION

Exercise is just physical activity and can be of various forms with each form having its own special benefits on our overall health. Strength training improves muscle endurance whereas Yoga improves flexibility and balance while Aerobics boosts our cardiovascular system. Each type of exercise is important in its own way, and doing all three types is the way to maximize fitness. Good physical health can work in unison with mental health to improve a person's overall quality of life.

1.1 Importance of Proper Form

When it comes to working out, quality is more important than quantity. It seems strength training and yoga both have an amazing set of benefits. This holds true as long as they are done correctly. If performed incorrectly, the results might be adverse. Everything we do physically becomes a neurological pattern [9]. A person when learning an exercise has a tendency to pick up an improper form which makes him/her prone to injury. Undue emphasis on muscles due to poor form leads to strains and sprains. Maintaining good form keeps one free from injury. It puts consistent tension on targeted muscles, leading to better results. Proper form directs our energy to the right set of muscles, thus helping us work out more efficiently.

1.2 Need of Virtual Coach

Previous section provided a good motivation to begin thinking about ways to ensure that we are exercising the right way. Personal trainers at the gym are definitely a good option, but post COVID, most of the gyms have shut down and many people have started preferring home workout over gyms. So, the next option is remote training, where usually people interact with their personal trainer by sending recorded workout sessions. This approach lacks real-time feedback. Also, people might not be comfortable sharing their videos. Enabling computers to perform the tasks of a personal trainer can overcome these problems. Steven Chen and Richard Yang [3] suggested two different approaches to correct exercise posture - a geometric approach and a heuristic machine learning based approach. A year later, Maybel Chhen Thar et al., [4] suggested a comparison based geometric model for Yoga Pose Assessment. Later, in 2020, Talal Alatiah and Chen Chen [8] went one step further and demonstrated a way to count repetitions in real time, again using a geometric approach. A major challenge in the geometric approach is that the perspective of the user frame must be the same as the reference frame, else the algorithm might fail.

2. RELATED WORKS

George Papandreou et al. [1] proposed the PersonLab model for Human Pose Estimation and Instance Segmentation. They used a box-free bottom-up (parts first) approach. The model consists of a convolutional neural network (to detect keypoints) and a part-induced geometric embedding descriptor (to map person pixels with its instance). Their system detects 17 face and body part key points for each person with a runtime that is independent of the number of people in the scene. The implementation of this system is now a part of the TensorFlow.js under the name PoseNet.

Zhe Cao et al. [2] also presented a real time approach to detect the 2D pose of multiple people in an image. The proposed model learned person-part mapping using Part Affinity Fields (PAFs). They, too, used a bottom-up approach and achieved high accuracy and real time performance, irrespective of the number of people in the image. This work

led to the release of OpenPose, an open-source real time system for pose estimation.

While OpenPose proves to be more accurate than PoseNet, it requires GPU powered devices to meet the real-time constraints. On the other hand, PoseNet has been built with the focus on lightweight devices and easy deployment on native and web applications for real-time pose estimation. PoseNet trades accuracy for speedy performance.

Steven Chen and Richard R. Yang [3] introduced Pose Trainer, an application that takes the user's exercise video, performs pose estimation with OpenPose, and gives feedback for improving the exercise form. They used poses and instructions of personal trainers and developed two models: 1) machine-learning based model, and 2) a heuristic model based on vector geometry. The model was trained using a dataset of over 100 videos spanning four common exercises. The application is supported only on GPU-enabled computers.

Maybel Chan Thar et al. [4] propose a Yoga pose assessment method using pose detection. The system first detects a Yoga pose using OpenPose, then, it calculates the difference of specific body angles between the trainer's pose and trainee's pose, and provides suggestions if it exceeds the threshold. The authors evaluated three people with three basic Yoga poses and concluded that the system found the incorrect parts of each pose successfully.

Y. Qi et al. [7] proposed a novel model for synthesizing dance moves from music sequences. They used OpenPose [2] to extract human pose estimation and used audio features like MFCCs [11] and trained a sequence to sequence [12] model with the above data and concluded that harmonious dance sequences can be generated using music melody.

3. PROPOSED WORK

In this paper, we propose Virtual Coach, an application that serves as a personal trainer. Virtual Coach is designed to train the user in three areas - weightlifting, yoga and cardio. Weightlifting section consists of strength-training exercises, with more focus on the ones involving dumbbells. While these strength exercises call for repetitions, yoga requires holding the body in a particular pose. Maintaining the proper form is vital for both weightlifting and yoga exercises. Virtual Coach was developed to help the user achieve this with the aim to prevent injuries. We use a sequence-to-vector model to evaluate the user. The user frames are captured in real-time and passed to a pose estimation model. The estimated poses are then input to the sequence-to-vector model which outputs a feedback tag (label) based on the evaluation. The detailed approach is explained in Strength Training [Section 4.1] under Methodology. Yoga poses, on the other hand, are assessed using vector geometry. We construct a geometric model which calculates the body angles and compares them with the ground truth reference pose features. The model outputs a set of labels identifying the good and bad points in the pose. Yoga Pose Assessment [Section 4.2] under Methodology lays out the working in greater depth. The third area of focus is cardio with dance. We integrate an auto-dance-generation feature for aerobic dances like Zumba that contribute to cardio workout. For this, we train a sequence-to sequence model that maps audio features to 2D dance pose coordinates which are then rendered on the screen for the user to follow along. The process has been further elaborated in Dance Generation [Section 4.3] under Methodology.

4. METHODOLOGY

4.1 Strength Training

The dataset for this model was created by performing pose estimation on publicly available workout videos. The samples were chosen such that they covered all possible mistakes in the exercise form as well the ones with perfect form. These video samples were then labelled with the appropriate classes such as CorrectForm, HalfROM, Swinging, and so on. We used PoseNet for pose estimation. Thus, the input vector contained 52 pose features - one representing the overall pose score and three for each of the 17 parts (score, x-coordinate and y-coordinate). We also added a one-hot encoded exercise name label as a part of the input features in order to distinguish between the same motion as correct or incorrect with respect to the exercise. For example, shrugging traps should be classified as correct while performing shrugs but it should be considered an incorrect movement while performing bicep curls. The output label was also represented as a one-hot encoded vector. The output labels were later mapped to appropriate feedback statements. We also augmented the data by vertical flipping. The final dataset contained 250 labelled samples.

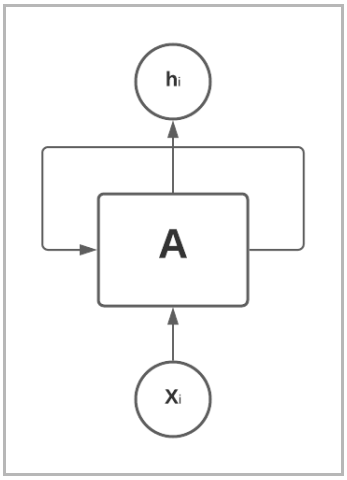

Figure-1: Single LSTM unit

The LSTMs consist of a special unit called as memory block which receives input through time in each time step, an

input vector is fed into LSTM, the output is computed according to:

$hi = fw(hi-1, xi)$

Figure 1 shows the general architecture of a single LSTM block. Each memory block in the original architecture contained an input gate and an output gate. The input gate controls the flow of input activations into the memory cell and the output gate controls the output flow of cell activations into the rest of the network. Just like we have Deep Neural Networks (DNN) which is an extension of NN we also have Deep LSTMs which have shown to give very promising results in many fields, in our case posture correction. In a Deep LSTM model simple LSTM models are stacked together. The inputs to the Deep LSTM model go through multiple non-linear layers as in DNNs, however the features from a given time instant are only processed by a single nonlinear layer before contributing the output for the next LSTM unit in the stack.

In our case we made use of one such Deep LSTM model with 4 LSTM layers and 1 dense layer on the top of it (initial inputs from LSTM layer). The general idea of the stacked Deep LSTM model is shown in Figure 2.

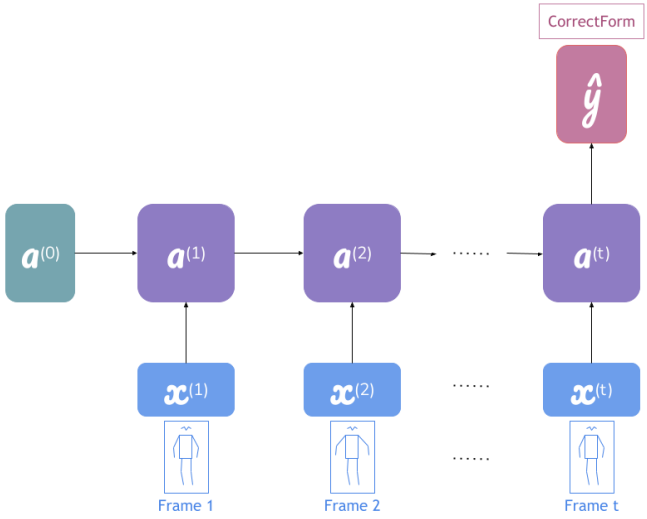

Figure-2: Deep LSTM model

The features extracted from the PoseNet model after preprocessing served as the input to the Deep LSTM model the preprocessing included splitting the data into train and validation and converting the output of the PoseNet model which were by default python lists to trainable Numpy array and one hot encoding the features. For training purpose tested we various models like RNN, GRU and Deep LSTMs with the LSTM model giving the highest accuracy among all. After hyperparameter tuning stacked LSTM (4 layers) with one dense Layer was trained with Adam Optimizer and the learning rate was 0.01 with a learning rate scheduler and the model was trained with 25 steps per epoch.

4.2 Yoga Pose Assessment

The yoga pose correction model uses a geometric approach similar to the one proposed by M. C. Thar et al. [4]. The body angle features extracted from the correct pose are prestored. These are used as reference to evaluate the user's pose. Since yoga requires the user to hold a body in a particular position, the user's pose is continuously compared with the reference pose.

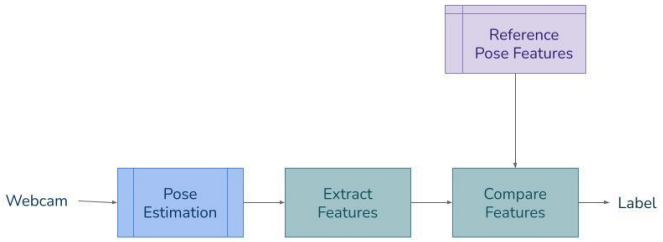

Figure-3: Workflow of Yoga Pose Assessment Model

First, pose estimation is performed in real-time on the user video captured via webcam. For every pose generated, body angles are calculated and compared against the reference pose. If the difference lies within a certain threshold, the pose is classified as correct and is rendered on the screen in green color. Otherwise, the parts where the difference exceeds the threshold are noted and highlighted in red. Appropriate instructions are also displayed as feedback based on the points where the user is going wrong.

4.3 Dance Generation

A huge volume of coordinates for Zumba dance [5] is needed to extract desirable features. To create a dataset with adequate and desirable data, we collected music and dance data as described below.

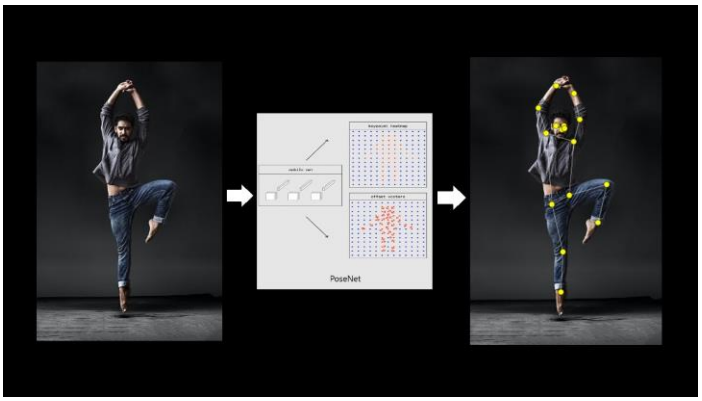

Figure-4: Pose Keypoints Extraction

For dance motion data, we collected various dance videos of professional Zumba [5] Dancers from various video streaming platforms like YouTube, Dailymotion, etc. and used PoseNet [6] to extract body coordinates (X and Y

coordinates on a 2-D plane) of these dancers as shown in Figure 4 and the coordinates' details are given in Table 1.

Table-1: Dance Pose Features

Index	Body Parts
1	Nose
2	Left Eye
3	Right Eye
4	Left Ear
5	Right Ear
6	Left Shoulder
7	Right Shoulder
8	Left Elbow
9	Right Elbow
10	Left Wrist
11	Right Wrist
12	Left Hip
13	Right Hip
14	Left Knee
15	Right Knee
16	Left Ankle
17	Right Ankle

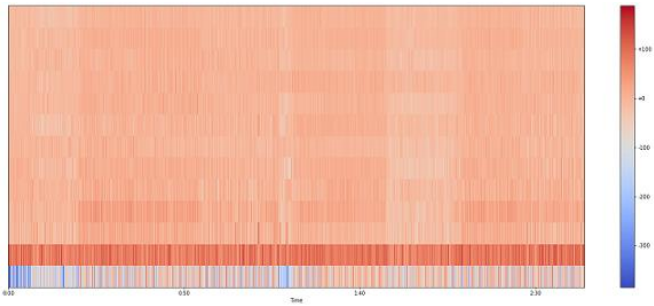

Figure-5: Mel-frequency cepstral coefficients (MFCC)

For audio data, the music on which these dancers were performing Zumba was used and an audio and music analysis package "librosa" [10] was used to extract audio features like Mel-frequency cepstral coefficients (MFCCs) [11] visualization of MFCCs as shown in Figure 5. In our case we have extracted 13 MFCCs. The steps for MFCCs extraction are: 1) take the Fourier transform (FFT); 2) Take the magnitude of FFT; 3) Extract MFCCs from output of the previous step; 4) Take log of the previously obtained features (i.e., convert it in dB(decibel)).

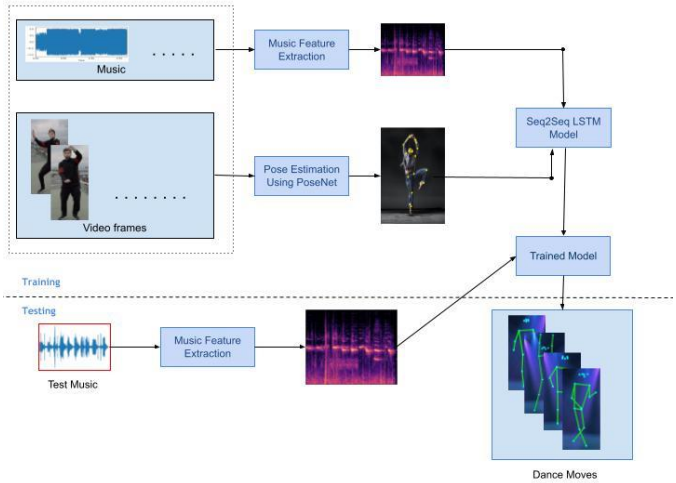

Figure-6: Workflow of dance generation model

The workflow of the dance generation model is shown in Figure 6. Dance generation model uses seq2seq [12] LSTM model which has 2 LSTM layers and then a dense layer. The music features extracted from the test music are passed to the model and the model generates human pose coordinates which are dance moves and these frames are then converted into a video file.

5. RESULTS

5.1 Strength Training

Figure-7: Evaluation of Bicep Curls

Results demonstrated that the model was able to evaluate the exercises even from different perspectives. Figure 7

shows the evaluation results of bicep curls. The model rightly classifies it as correct.

5.2 Yoga Pose Assessment

The geometric approach proved to be very effective. Figure 8 shows the performance of the system when evaluated on Tadasana (or the mountain pose). It can be observed that the model correctly marks the areas contributing to bad pose and provides proper feedback. It also encourages the user when the pose is correct.

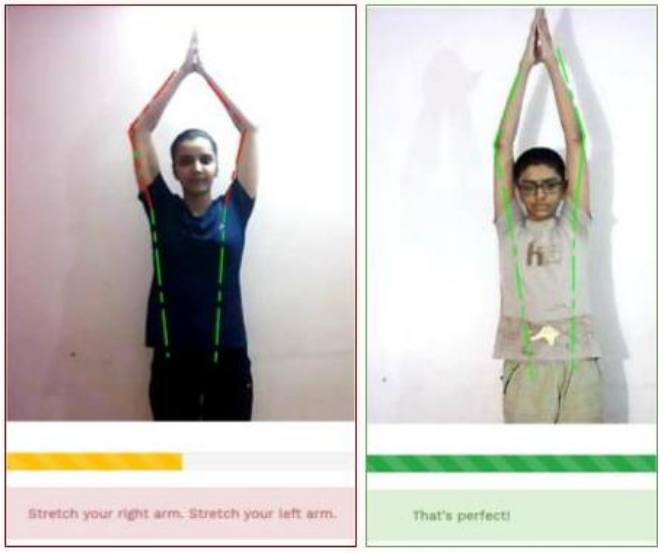

Figure-8: Assessment of Tadasana

In Figure 9, we can see the assessment of Vrikshasana (the tree pose). The model guides the user on the left with further instructions. It also recognizes the correct pose successfully.

Figure-9: Assessment of Vrikshasana

Thus, the results prove that the model was able to correctly identify areas that needed improvement and generated appropriate instructions for taking corrective actions.

5.3 Dance Generation

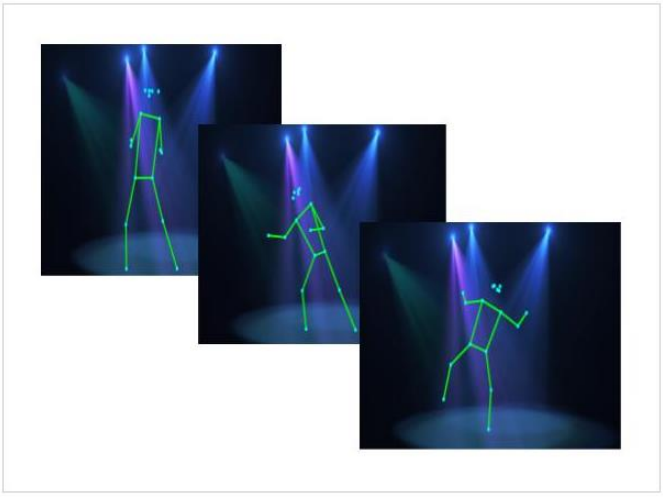

Figure-10: Generated 2D Dance Moves

The generated human pose keypoints by the dance generation model were harmonious and thus the user can follow along the skeleton to perform Zumba. Figure 10 shows some dance moves generated by the model.

6. CONCLUSION

In this paper, we put forth the idea of Virtual Coach - a digital fitness trainer, who monitors the user's exercise form and suggests improvements, if necessary, thus making the workout more efficient. Virtual Coach also introduced a unique feature that generates dance moves in real-time given any music for aerobic dance workouts like Zumba with the help of sequence-to-sequence model. The exercise form assessment for weightlifting was done using a sequence-to-vector model, while that of Yoga was done with vector geometry. The system was evaluated by three people and results were recorded. All three models gave acceptable results. Overall, the application works fine and serves the purpose.

7. CHALLENGES

Pose Estimation is the backbone of Virtual Coach. Virtual Coach uses PoseNet for pose estimation. Flickering poses were a big hurdle in making sequential models learn the right mappings. Using OpenPose might improve the accuracy but the challenging part here is that OpenPose requires a GPU to perform pose estimation within milliseconds. Virtual Coach being a real time application cannot rely on slow CPUs for processing. One solution could be to buy compute nodes from some cloud provider and let the GPU-enabled servers take care of all the processing. Even though using cloud computing for setting up GPU-enabled servers will solve the problem of low computing speed of CPUs, the lag for real

time data upload and download still remains to be addressed.

8. FUTURE WORK

Virtual Coach is an application in the health domain and health is incomplete without diet. Thus, recommending diet will be a major integration in the software. We can store the nutritional preferences and physique information about the user which can then be used in recommending personalized workout and diet plans that suit the needs of the user. The recommendations can be further personalized by taking into account the age, environment, lifestyle and medical history of the user.

REFERENCES

[1] George Papandreou, Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, Kevin Murphy, "PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 269-286, arXiv:1803.08225.

[2] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2929257.

[3] Steven Chen and Richard R. Yang, "Pose Trainer: Correcting Exercise Posture using Pose Estimation," 2020, arXiv:2006.11718.

[4] M. C. Thar, K. Z. N. Winn and N. Funabiki, "A Proposal of Yoga Pose Assessment Method Using Pose Detection for Self-Learning," 2019 International Conference on Advanced Information Technologies (ICAIT), 2019, pp. 137-142, doi:10.1109/AITC.2019.8920892.

[5] Zumba https://en.wikipedia.org/wiki/Zumba.

[6] Real-time Human Pose Estimation in the Browser with TensorFlow.js https://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html

[7] Y. Qi, Y. Liu and Q. Sun, "Music-Driven Dance Generation," in IEEE Access, vol. 7, pp. 166540-166550, 2019, doi: 10.1109/ACCESS.2019.2953698.

[8] Talal Alatiah and Chen Chen, "Recognizing Exercises and Counting Repetitions in Real Time," 2020, arXiv:2005.03194

[9] Why form is important in working out https://www.fitness19.com/why-form-is-important-in-working-out/

[10] librosa: a python package for music and audio analysis. https://librosa.org/doc/latest/index.html

[11] J. H. Jensen, M. G. Christensen, M. N. Murthi and S. H. Jensen, "Evaluation of MFCC estimation techniques for music similarity," 2006 14th European Signal Processing Conference (EUSIPCO), 2006, pp. 1-5.

[12] Ilya Sutskever, Oriol Vinyals and Quoc V. Le, "Sequence to Sequence Learning with Neural Networks," In Advances in Neural Information Processing Systems (NIPS), 2014, arXiv:1409.3215