International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Yoga AI Assistant with ReactJS

Rajiv Gandhi Institute of Technology, Bangalore-560032

Gayathri S Bhat, Saritha IL, Hariharan, Honnur Swamy

Department of CSE,
Under the guidance of Dr.Arudra(HOD)

Abstract - Yoga is beneficial for maintaining physical and mental fitness. And these days, technology is helping yoga find a new companion as online yoga becomes more and more common, making it easy to access classes and instructions. TensorFlow's pose net, a tool for pose detection, is used by the AI yoga web app. Pre-trained models for identifying human poses are immediately accessible in the browser using posenet.js, a toolkit for estimating yoga poses based on body count rate determined by the pose. A person can receive feedback on the accuracy of their pose and keep track of their fitness goal in the application using real-time camera data.

1.INTRODUCTION

The model is trained and deployed using Tensoeflow.js. Pose Net is an open-source model for online pose estimation that comes from Tensorflow.js. Pose estimation is a computer vision technique that identifies a person's pose in pictures and videos. What it can do is 1. Proper Body Position 2. Strengthen acrobatic and athletic poses (Tells you how flexible you were) 3. Workout/Yoga (Finished pose/movement tracked). We can begin to recognize some fundamental poses using Pose Net and a few JavaScript conditions. The system will be able to give the user feedback on the posture's accuracy based on the user's pose as shown in the video frames. We have seen in our family that our elderly suffer from flexibility issues and broken muscles. Because people neglected to pay attention to their body posture while exercising or doing yoga, they are the ones who suffer from this. To address this problem, we have developed this innovative solution. We created an AI yoga assistant after observing all the challenges they faced. Our assistant will assist users in keeping proper posture while engaging in yoga and fitness. We have seen in our family that our elderly suffer from flexibility issues and broken muscles. Artificial intelligence of computer vision enables computers and systems to extract useful information from digital photos, videos, and other visual inputs and to conduct actions or offer recommendations in response to that information. If AI gives computers the ability to think, computer vision gives them the ability to see, observe, and comprehend. Human vision has an advantage over computer vision in that it has been around longer. With a lifetime of context, human sight has the advantage of learning how to distinguish between things, determine their distance from the viewer, determine whether they are moving, and determine whether an image is correct. Computer vision teaches machines to carry out these tasks, but it must do so with cameras, data, and algorithms in much less time than retinas and visual cortex. A system trained to inspect items or monitor a production asset can swiftly outperform humans since it can examine thousands of products or processes per minute while spotting imperceptible flaws or problems. A lot of data is required for computer vision. It repeatedly executes analyses of the data until it can distinguish between things and recognize images. For instance, a computer must be fed a massive amount of photos and data in order to teach it to recognize it. Automatic self-training techniques for athletes can improve their performance and lower their risk of injury. For analyzing exercise-related activities including football player rankings, handball strikes, volleyball, sprinting, jumping, and other athletic activities, many researchers have devised computerized methods. A features-based approach to developing a self-training system that could identify yoga poses. It uses a Kinect to produce a body map by capturing the user's body contour. A star skeleton was used for rapid skeletonization in order to generate a descriptor for the human posture. A yoga recognition system based on a Kinect and the AdaBoost classification is suggested for six asanas, with a

94.78 accuracy score. However, they are using a depth sensor-based camera, which is often unavailable. In order to identify Indian traditional dance and yoga poses from images, Mohanty et al. created an image identification method using convolutional neural network (CNN) and stacked autoencoder (SAE) methodologies. Yet, they only ever evaluated their skill on still images. The conventional skeletonization process has been replaced by deep learning-based technologies ever since Toshev et al. introduced Deep Pose. Leading the movement away from conventional methods and towards deep network-based methods is Deep Pose. Using regressors built on deep neural networks, it directly regresses on joint coordinates. It forecasts a person's movements and the location of concealed bodily components. Yet, localization is a challenge for their strategy.

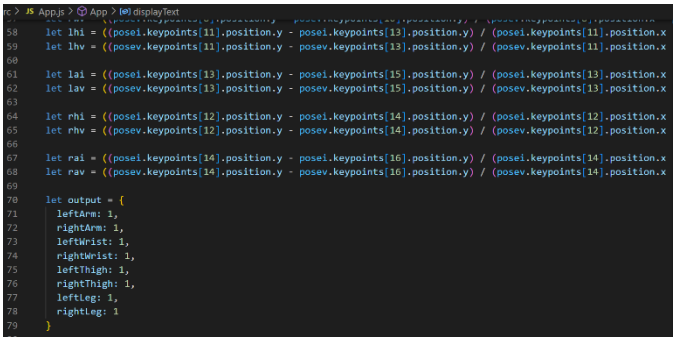
.

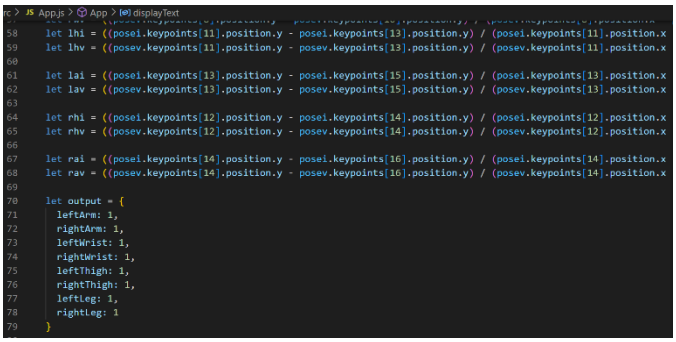
2. Body of Paper

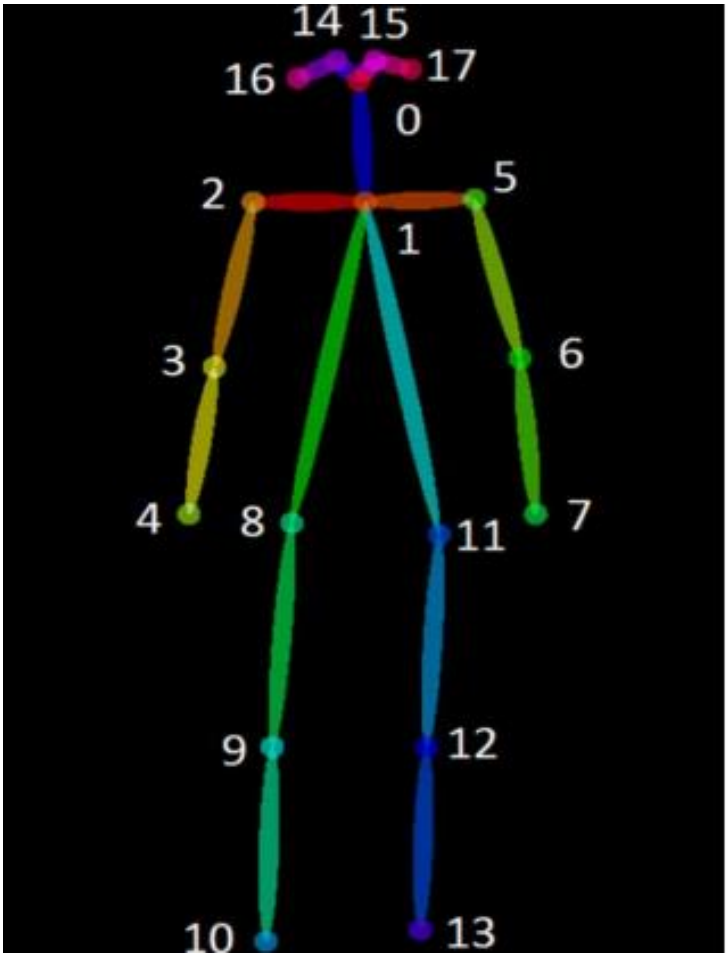
Robotics and computer engineering are just two areas where human activity recognition has been used. Randomized trees (also known as random forests) are used in references to detect human activity using sensors. Reference identifies human activity using hidden Markov models and identifiable body components. The recognition of six domestic activities using this method had a 97.16 percent accuracy rate. For monitoring purposes, smart houses use this technique. Wearable sensors that can detect sounds are used, and they have a 96.9 percent accuracy rate when used for human activity detection. Much effort has gone into creating automated systems that can evaluate yoga as well as sports like basketball and cycling. Using kinetic sensors and an AdaBoost classifier, an automated project for yoga stance detection achieved 94.8% accuracy. For three yoga positions, a different approach shown in reference obtained 82.84% accuracy. System classified yoga positions using deep learning algorithms. While deep learning analyses data and extracts features, typical machine learning models call for extracted features and engineering. Star skeleton computation is used to create a self-teaching system for the yoga position. Kinect is used to derive the body contour from the user's body map, and an accuracy of 99.33 percent was attained. It extracted the human pose from the pressure sensor using hash-based learning. Yoga poses are classified using pose estimation in OpenPose and a hybrid CNN with LSTM model that also includes feature extraction. Additionally, it contrasted machine learning models with deep learning models and simple CNN models with a hybrid model. The classification score and confusion matrix are employed as evaluation metrics. Test accuracy was 0.9319 for SM, 0.9858 for CNN, and 0.9938 for the hybrid CNN with LSTM model. There are numerous key point detection techniques, including PoseNet, OpenPose, and PifPaf. Key points are obtained using CNN-based architecture and the CMUinvented Open Pose. VGG-19 is used by OpenPose to extract features from photos. The first branch found 18 confidence maps (initial layers). Predictions are made by using the second branch. For estimating the relationship between body parts, use the second branch. PoseNet and OpenPose, which can extract human stance, are comparable. The confidence levels for each of these essential criteria are listed, with 1 representing the highest and 0 the lowest. PoseNet does not depend on the size of the photographs; the pose is still recovered even when the images are downscaled. The final pose is regressed using the regressor after the final pose has been encoded, localised, and converted to a localization feature vector using the +e encoder. PifPaf bases its human posture extraction on a bottom-up methodology. Body parts are localised using a Part Intensity Field, and they are associated with one another using a Part Association Field, which together make up the full body posture. ResNet is the architecture that is being used. These models use 18 key points, and each key point's input size is 36 x and y coordinates. If features are taken from these focal points, models will be trained with greater precision. Twelve separate joints are employed in the project to extract 12 features, or angles, which are then used as model inputs. The relationship between human activities and angle motion sequences can be seen in earlier techniques, which used joint angles as characteristics. These angle characteristics have more information than key points since they are scalable. The angles between the elbows, shoulders, knees, and crotch are proven to give more information for 3D human activity detection in reference [25]. Angle pairs for the hip bones are added in reference, are employed as features in the references so that features can be resized. Hip and knee angles are the major characteristics used in reference. The shoulders and knees create angles at the hip joints, and the hips and ankles create angles at the knees. These features provide more information than key points because the angles recovered at any distance from the camera will be the same, unlike key points which are not scaled. Angles are determined with relation to a reference

3 METHODOLGY

Yoga asanas performed by the user are automatically recognised from both current and archived video. The system performs the subsequent actions: The first source of information is a live camera feed or a previously archived video. After that, Posenet locates the joint. The identified points are sent to the KNN classification model. Thirdly, the pre-trained model is utilised by the application. One can generate 12 joints from these 13 vectors by deleting the vectors. Consider that the body point at the neck is and the body point at the right shoulder is (x1, y1) (x2, y2). Their vectors are x1i + y1j for the neck and x2i + y2j for the right shoulder. Take the neck vector, which is equal to (x2 x1)i + (y2 y1)j, away from the shoulder vector. To get a vector for the joint neck and right shoulder, as shown in Figure 1. Nevertheless, because the origin in images is at the upper left corner rather than the lower left corner, 1 should be multiplied by (y2, y1). The vector of the joint is therefore (x2, x1)i + ((y2, y1)j. With this procedure, 12 vectors for 12 joints are produced, along with the angles that each joint is generating.







4 CONCLUSION

The expected result of the project is to detect Yoga poses and track if the user is able to hold on to the pose for a specified time interval. This would be helpful for the user to practice. The timer can be set by the user which provides both beginner and advanced Yoga practitioners to enhance the skills according to their ability. The counter stops when the user fails to remain in a particular Yoga pose and restarts only when the user gets back to the correct pose. The project can be further enhanced by adding more number of Yoga poses and interactive sound responses to the application. Neural networks (MLP) are built using 3 types of layers, namely, input layer, hidden layers, and output layer. +ere can be any number of hidden layers based on the complexity of training data. If hidden layers are few, the model may underfit training data, and if they are more, the model may overfit. MLP is a fully connected neural network, that is, every node is connected to every other node in consecutive layers in the neural network. Generally, these networks are utilized for supervised training where for every input data, there is a corresponding output label or class. Both training and validation datasets have many ups and downs in accuracy till the 6900 epoch and attained an accuracy of 0.9958 at the 6900 epoch. From 6900 to 10000 epochs, loss of training and validation decreased gradually, which results in the training model classifying with high confidence. From epoch 0 to 10000, the loss of validation and training datasets decreased gradually. Both training and validation datasets have many ups and downs in accuracy till the 6900 epoch and attained an accuracy of 0.9958 at the 6900 epoch. From 6900 to 10000 epochs, loss of training and validation decreased gradually, which results in the training model classifying with high confidence. From epoch 0 to 10000, the loss of validation and training datasets decreased gradually

References

[1] Y. Wang, J. F. Doherty, and R. E. Van Dyck, “Moving object tracking in video,” in is are Applied Imagery Pattern Recognition Workshop, 2000, vol. 2000-January, pp. 101.

[2] B. Tian, Q. Yao, Y. Gu, K. Wang, and Y. Li, “Video processing techniques for traffic flow mon A survey,” in IEEE Conference on Intelligent Trans human catenation Systems is are, Proc , IT 2011, pp. 1103– 1108.

[3] J. M. B. Oñate, D. J. M. Chipantasi, and N. D. R. V. Erazo,“Tracking objects using Artificial Neural Networks and wireless connection for robotics,” J. Telecommunication. Electron. Computer. Eng .,vol. 9, no. 1–3, pp. 161–164, 2017.

[4] M. Brown, J. Funke, S. Erlien, and J. C. Gerdes, “Safe driving envelopes for path tracking in autonomous vehicles,” Control Eng. Pract., vol. 61, pp. 307–316, 2017.

[5] V. A. Laurense, J. Y. Goh, and J. C. Gerdes, “Pat - tracking for autonomous vehicles at the limit of friction,” inProceedings of the American Control Conference, 2017, pp. 5586–5591.

[6] S. Walker et al., “Systems and methods for localizing, tracking and/or controlling medical instruments,”2017.

[7] M. Buric, M. Pobar, and M. Ivasic-Kos, “An overview of action recognition in videos,”2017 40th Int. Conv. Inf. Communication. Technol. Electron. Micro electron. MIPRO 2017 - Proc., pp.

[8] "Human joint angle estimation and gesture identification for assistive robotic vision," in Proceedings of the European Conference on Computer Vision, Springer, Amsterdam, +e Netherlands, October 2016, pp. 415-431. A. Guler, N. Kardaris, S. Chandra, et al.

[9] "Real-time Yoga recognition using deep learning," Neural Computing & Applications, vol. 31, no. 12, pp. 9349-9361, 2019. S. K. Yadav, A. Singh, A. Gupta, and J. L. Raheja.

[10] Multimedia Tools and Applications, vol. 78, no. 20, pp. 29357-29377, 2019, P. Szczuko, "Deep neural networks for human position estimation from a very low-resolution depth image."