International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 61-65
Published Online November 2022 in IJEAST (http://www.ijeast.com)

REAL TIME YOGA POSE DETECTION USING DEEPLEARNING: A REVIEW

Khushi Sidana
Computer Science
Narsee Monjee Institute of Management Studies
Mumbai, India

Abstract—**With the increase in the number of yoga practitioners every year, the risk of injuries as a result of incorrect yoga postures has also increased. A self-training model that can evaluate the posture of individuals is the optimal solution for this issue. This objective can be attained with the aid of computer vision and deep learning. A model that can detect theyoga pose performed by an individual, evaluate it in comparison to the pose performed by an expert, and provide the individual with instructive feedback would be an effective solution to this problem. Recently, numerous researchers have conducted experiments on the detection and performance of yoga poses in real time. This paper discusses the methods undertaken in brief and compares the tools and algorithms they used for conducting pose estimation, pose detection as well as pose assessment. Itdiscusses the accuracy, precision, and similarity of pose classification obtained by the researchers and the future scope of the research.**

Keywords—**computer vision, deep learning, human pose detection**

I. INTRODUCTION

Yoga is an ancient practice, originated in India, practiced with the objective of bringing together the soul, mind and body. Yoga encourages a balance between these three pivotal parts of human's being. Yoga has many advantages and with time, humans all over the world are getting more educated andaware of the health benefits yoga provides. It helps increase and build muscle strength along with increasing flexibility andposture. Yoga also reduces chronic pain and helps reduce the chances of health disease. A major part of yoga are medication exercises which along with improving respiration, help increase attention and focus as well. Yoga is also very successful in reducing stress and helping individuals stay calm.Talking about the impact it has on our mind, yoga promotes self healing and personal power. It helps us remove the restrictive mental blocks we have in our mind and view the world more positively. Owing to the plethora of benefits yoga provides, the number of people performing yoga is increasing at an exponential rate. In the last five years, yoga practitioners have increased by around 5% and currently approximately 300 million people regularly practice yoga globally [1]

Human pose estimation is a task based on computer vision that is being widely applied for a vast amount of applications in recent times. Human pose estimation is basically a technique for identifying the joints or key points in the body. The key point is given as a set of co-ordinates that identifies the given location of the body. Human pose estimation can be of two forms, them being 2D pose detection as well as 3D pose detection. Human pose estimation is essential to help understand the body language and thereby help computers perform activity recognition. This field of human pose estimation will have a huge impact on quite a few Real-time applications. It will largely impact the fields of robotics and autonomous driving. It could also be used for augmented reality and motion tracking, especially for games. Along with pose detection, pose grading involves grading the pose performed by the individual in contrast to the target pose or thepose performed by the expert. There are many methods to perform pose estimation including skeleton-based model, contour-based model and volume-based model that we will look at in detail.

Since the number of individuals performing yoga are using exponentially, the need for them to perform safe and accurate yoga is also rising. Any kind of misalignment or incorrect pose can cause a wide range of problems ranging from acute pains to chronic problems. individuals performing incorrect poses are prone to fractures, sprains, joint dislocations, nerve damage and even stroke. A self-trained model that will help individuals correct their poses in accordance to the similarity of the poses with that of the expertcould help reduce these problems caused by incorrect poses and could be beneficial for the individuals. Considering this, real time yoga pose estimation and grading can prove to be very beneficial.

II. SURVEY ON METHODOLOGIES USED

A. Pose Estimation

In the survey, many pose assessment methods for 2D yoga poses are observed. The key points can be found using a variety of techniques, including as OpenPose, MediaPipe, TensorFlow, and others. In essence, this stage entails

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 61-65
Published Online November 2022 in IJEAST (http://www.ijeast.com)

estimating poses on both the student or person doing yoga in real time as well as the yoga expert in the static image. We need to record the video using a camera and extract the frames from the recording in order to estimate the pose in real time. There are several other strategies that have been employed in research on estimating postures, which we will briefly cover.

H-T. Chen et al. suggested capturing the video for real time data using a Kinect in the paper [2] published in 2014. They used the OpenNI library to extract the body map from the Kinect after positioning it about 200 cm away from the subject. The extracted body contour is then obtained by smoothing the body map, which is then utilised to construct the star skeleton. For upcoming comparisons, we compute the star skeleton of the target poses utilising their static structures as well. This Kinect based approach proved to be not so convenient since Kinect sensors are expensive and are not readily accessible. Considering that, focus was more on the usage of deep learning for human pose estimation which can be seen in the future studies. Rutuja Gajbhiye et al. proposed to perform pose extraction both online in real-time and offline byextracting the key points in the frame using the OpenPose library in the paper [3] published in 2022. OpenPose captures the video and processes it frame by frame, extracting the key points and saving its corresponding output in JSON format. This JSON data retrieved is then stored in a NumPy array in a 45-frame sequence. The data is divided in 60-20-20 where 60%of the data is utilised to train the model, 20% is used for testing, and 20% is used to validate the data. Each sequence then contains the 18 points with two co-ordinated detected using OpenPose.In the study[4] released in 2019, Santosh K. Yadav et al. carried out pose extraction utilising OpenPose once more. The 18 keypoints identified by OpenPose were retrieved from the output in JSON format. In order to execute this as efficiently as possible, they operated at 3FPS while using the default resolution. They partitioned the dataset into sections that were utilised for training, testing, and validating in a 60-20-20 manner. K.Z.Winn et al. performed pose extraction by using OpenPose once again which uses part affinity and convolutional neural networks to detect the points in the human body[5].

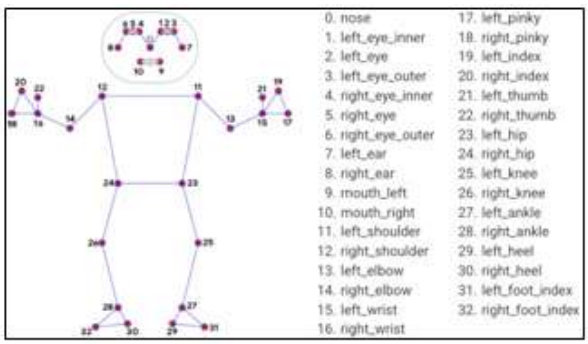

Figure 1. Key points located using MediaPipe

Yubin Wu et al. proposed a novel approach of employing contrastive examples for the yoga positions in another research [6] that was released in 2022. They came up with two differenttypes of sets: coarse triplet sets, which include an anchor, a positive and a negative illustration from a different category ofposes, and fine triplet examples, which also include an anchor,a positive and a negative illustration from the same category of poses but with various pose characteristics. Here, they employed Media Pipe, another method for locating the key points of the human body, to extract the skeleton. They used MediaPipe to extract the key points using the learning model

BlazePose. With the aid of Media Pipe, we are able to obtain the 33 essential locations on the human body, each with two coordinates, so that the image has (2,33) coordinate data values. These points can be seen in Figure 1.

Radha Tawar et al also used MediaPipe to extract the key points from and create a skeleton using these points [7]. They performed this without using any contrastive examples though. In their study [8], D. Swain et al. employed MediaPipeto identify the critical points, and then they used those points to compute angles to carry out posture assessment by angle dissimilarity.

In a unique strategy, [9] Chhaihuoy Long et al. used transfer learning in place of key point identification. To perform transfer learning, they used six pre-trained models with weights that were trained separately to extract the featuresof the yoga pose. Transfer learning basically helps to predict the target database with the help of two layers that are feature extraction and classification layer.

B. Pose Detection and Assessment (Accuracy)

The next step is to train the model to reliably detect and anticipate the poses after the pose estimation stage has identified the key points. For pose detection, a variety of machine learning, deep learning, and hybrid algorithms have been utilised to achieve the most accurate results. These approaches are briefly covered in the survey that was conducted. Using these approaches, the accuracy is calculated using the notation given below:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Accuracy can also be calculated using the confusion matrix which is a method used to summarise the performance of the model or algorithm. Figure 1 shows us what the confusion matrix looks like which includes True positive, True negative, False positive and false negative values.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 61-65
Published Online November 2022 in IJEAST (http://www.ijeast.com)

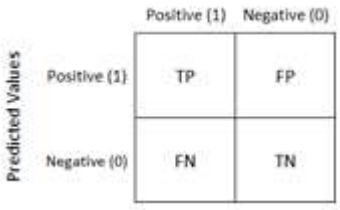

Figure 2. Confusion Matrix

Using the confusion matrix, accuracy can be given by the following formula or notation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Rutuja Gajbhiye et al. introduced in their study [3] a new hybrid strategy that combines deep learning and machine learning classifiers. In this method, the support vector machine (SVM) is first employed, which employs machine learning expertise to improve the overall performance of machine learning algorithms. The skeletal structure obtained with the aid of key pose detection of the individual and of the target pose is then fed as inputs to convolutional Neural Networks (CNN) in order to gain the similarity score and obtain the accuracy. They attained a training accuracy of 99.53 %, a testing accuracy of 93.19 %, and a validation accuracy of 97.6
% using the suggested method. 6992 frames out of a total of 17865 were misclassified. In order to conduct a comparison study, they detected poses using solely CNN. They attained training accuracy of 98.41%, testing accuracy of 98.68%, and validation accuracy of 99.10% with this method. Here, just 93 frames were incorrectly categorised. In spite of the fact that CNN achieves greater accuracy than the hybrid model, we can deduce that there is some overfitting because the model loss curve demonstrates an increase in validation loss and a decrease in testing loss.

Yubin Wu et al. proposed two baseline methods [6] for conducting the experiment and predicting the accuracy of yoga poses. The baseline approaches are centred on the proposed system, which aims to maximise the similarity between yoga poses through a contrastive loss in the latent feature space. For this, they employ a neural network encoder that extracts representation vectors from the contrastive data examples that are input. The coarse sets and respective skeleton points are then input into the shared weight encoder. Using this, we obtain a triplet contrastive loss and then conduct a comparisonof feature similarity. The two baseline approaches and the proposed methodology are compared to ensure the highestdegree of precision possible. Using only coarse contrastive examples, the proposed method achieves an accuracy of 77.6%, whereas accuracy increases to 83.21 % when using both fine and coarse contrastive sets.

Radha Talkwar et al. proposed a system [7] to perform pose assessment by using angle calculation and detection after getting the skeleton detection using MediaPipe. The angle is calculated using the formula :

$$\tan(x + y) = \frac{\tan x + \tan y}{(1 - \tan x * \tan y)}$$

Using this formula, the angles of various joints were calculated. The pose's labels are initialised. Once the extracted angles of an individual's skeleton are compared with those of experts, audio output is provided in the form of text-to-speech feedback. Additionally, a GUI has been developed to display the output.

Santosh K. Yadav et al. proposed using a hybrid system
[5] comprised of CNN and LSTM to predict yoga poses. This method was used to identify six yoga poses. Once the CNN layer has extracted the features using the 2D coordinates of the key points, the LSTM layer analyses the variations in these features across successive frames. The softmax layer then computes the likelihood of each yoga pose in the frame. Pythonwas used to programme this model using the Keras Sequential API. They achieved a training accuracy of 99.34% and a testing accuracy of 99.04% using this method. They achieved atesting accuracy of 99.38 % and a real-time accuracy of 98.92 percent by polling 45 frames to reduce data loss duringtransition.

After receiving the features from the pre-trained model, C. Long et al. [9] used deep neural networks to predict 14 classes of yoga poses using the features from the pre-trained model. Softmax layer was used by the network employed in the study to perform multi-class classification. In addition, 100 epochs of categorical cross-entropy loss function were utilised in the training process. They utilised numerous programming languages, including TensorFlow, Keras, OpenCV, Python, etc. They determined the accuracy of the predictions in terms of specificity and sensitivity using the performance matrix. In addition to developing predictive models, they also developed a coaching feedback system. They performed angle calculations for each human joint, including the left/right elbow, the left/right knee, etc. The prediction is compared to the predicted pose of the yoga expert, and the key point selection is used to calculate the specific angles. A threshold is established for the allowable difference in angle between the individual and the instructor; if the value of the difference in angles exceeds this range, the user receives feedback to help them improve. The overall accuracy of the model was 98.43%. During the classification performance, warrior 2 exhibited the highest overall accuracy of 98.43%.

Dehabrata Swain et al. again used the hybrid CNN+ LSTM
[8] system to classify, predict and assess the images. They usedthe time distributed system along CNN using 16 filters

and a window size of 3. The model is trained for a total of 50 epochs since the growth becomes study after a long exponential growth then. After each epoch, the accuracy is checked to ensure we get the best possible accuracy. After calculating angle similarity with the cosine similarity function, they also performed pose correction by providing the appropriate feedback. A maximum threshold is also provided, indicating the maximum permissible deviation from the pose performed by the expert. When this threshold is surpassed, feedback is provided. In addition, they performed prediction by polling 45 frames and considering the Mode of the independent outputs. Using this technique, they attained a test accuracy of 99.53 percent, as well as precision and recall values of 0.9866 and 0.9869, respectively.

Khine Z. Winn et al. developed a yoga pose evaluation method [5] in which they rated the individual's performance as "perfect," "good," "fair," and "not good." After constructing the human skeleton with the aid of Open Pose, the angles between all joints were calculated. After calculating the angles of both the instructor and the student, the difference was computed using the formula:

Difference = | Instructor angle - learner angle |

These obtained angle differences were then saved in the form of an array based on the sequence of joints displaying the deviation. To make it easier for the user to comprehend, especially since he or she is a self-learner, a colour system was implemented. Each joint's angle difference was assigned a corresponding colour value. Red represented a high difference value, indicating that the pose around the joint was good, whereas green represented a low difference value, indicating that the pose was not good. This system will aid the user in early comprehension and self-correction. The system also evaluates and labels the pose as "perfect," "good," "fair," "not good," or "bad." The result is calculated by dividing the total angle difference by the total number of joints. With the range function, this value is mapped. They demonstrated the effectiveness of their system by applying it to three individuals of various ages, shapes, and body types.

H-T. Chen et al. carried out yoga pose detection using the star skeleton method [2]. They used a body contour to represent the physical posture and then computed the star skeleton structure of the skeleton we obtained. This is the concept of connecting the body's centroid to its gross extremities, for which the distance is calculated in a clockwise direction. This distance function is then smoothed with a Gaussian smoothing filter in order to identify the extremities. The feature is then presented as a set of vectors and is normalised to ensure vectors of uniform size, as feature sizes can vary. This method presented them with an error in which mismatched vectors are sometimes neglected, which can result in inaccurate predictions. They overcame this difficulty with the assistance of a penalty mechanism for mismatched vectors, as opposed to ignoring them. Overall, the accuracy of the proposed star skeleton system was quite high at 99.33%.

III. CONCLUSION AND FUTURE WORKS.

As we have discussed, systems for the detection and evaluation of yoga poses based on deep learning have been developed in the past using a variety of different techniques. The majority of these techniques have achieved an accuracy between 97% and 99.9%. Pose estimation strategies such as OpenPose, MediaPipe, and Transfer learning have been extensively discussed. Pose Detection has been refined using CNN, SVM, a CNN-LSTM hybrid model, deep neural networks, and a novel concept of star skeleton structure. Multiple studies have focused on the development of self-training and evaluation systems after the accurate and efficient detection of yoga poses. They did so by utilising the similarity of the calculated angles at the joints. In order for the users to self-learn, they have employed novel and creative methods for displaying the results. While they have successfully completed these experiments, the future objective is to apply these methods to a larger dataset containing more yoga asanas. It also involves decreasing the response's time complexity. This research can also contribute to the expansion of this model's application to the detection of other human activities, such as those related to sports, healthcare, etc.

IV. REFERENCES

[1]. Yoga Earth, https://yogaearth.com/yoga-research/yoga-statistics/

[2]. Chen, HT., He, YZ., Hsu, CC., Chou, CL., Lee, SY., Lin, BS.P. (2014). Yoga Posture Recognition for Self-training. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds) MultiMedia Modeling. MMM 2014. Lecture Notes in Computer Science, vol 8325. Springer, Cham.

[3]. Rutuja Gajbhiye; Snehal Jarag; Pooja Gaikwad; Shweta Koparde (2022).AI Human Pose Estimation: Yoga Pose Detection and Correction,International Journal of Innovative Science and Research Technology.

[4]. Yadav, Santosh & Singh, Amitojdeep & Gupta, Abhishek & Raheja, Jagdish. (2019). Real-time Yoga recognition using deep learning. Neural Computing and Applications. 31. https://link.springer.com/article/ 10.1007/s00521-019. 10.1007/s00521-019-04232-7.

[5]. 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..

[6]. Wu Y, Lin Q, Yang M, Liu J, Tian J, Kapil D, Vanderbloemen L. A Computer Vision-Based Yoga Pose Grading Approach Using Contrastive Skeleton Feature Representations. Healthcare (Basel). 2021 Dec 25;10(1):36. doi: 10.3390/healthcare10010036. PMID: 35052200; PMCID: PMC8775687.

[7]. Radha Tawar , Sujata Jagtap, Darshan Hirve, Tejas Gundgal, Dr. Namita Kale, Realtime Yoga Pose Detection, International Research Journal of Modernization in Engineering Technology and Science

[8]. Dehabrata Swain, Santosh Santapathy, Pramoda Patro, Aditya Kumar Sahu (2022). Yoga Pose Monitoring using Deep Learning, Research Square.

[9]. Long C, Jo E, Nam Y. Development of a yoga posture coaching system using an interactive display based on transfer learning. J Supercomput. 2022;78(4):5269-5284. doi: 10.1007/s11227-021-04076-w. Epub 2021 Sep 20. PMID: 34566258; PMCID: PMC8451169.