YOGA POSE DETECTION USING OPENPOSE AND MEDIAPIPE

GEETHA M[1], VISHWAS B M[2]

[1]Assistant Professor , Department of Master of Computer Applications , BIET College , Davangere
[2]Student , Department of MCA , BIET College , Davangere

---***---

Abstract - ***An abstract i.e. idea behind this yoga pose detection project using deep learning or neural network learning is that yoga popularity is increasing day by day because of its benefits. Doing yoga helps us physically, mentally as well as spiritually. Because of this many people nowadays are doing it regularly. The main idea of this project is to help the people to recognize which yoga pose they are doing with the help of this detection technique. Yoga which involves 8 rungs and limbs of it, which includes Yama, Niyama, Asana, Pranayama, Dharana, Dhyana and Samadhi. To easily help people understand which pose they are performing via images, video recording by classifying it, we are implementing this project because of this people will incline towards doing more as they will get help to identify which pose they are doing very easily.***

***Key Words*: CNN, Deep learning, Feature Extraction,Yoga Pose Detection, openPose , mediaPipe**

I. INTRODUCTION

The word yoga was derived from the Sanskrit word ‘yuj’ which means ‘to join’ or ‘to unite’. This union is not, merely, about your nose touching your knees as you bend to touch your toes! The union referred to is that of your mind with your body. You integrating with your surroundings and nature. And, finally, your individual consciousness with the universal consciousness[2].

Several Thousand years ago, on the banks of the lake Kantisarovar in the Himalayas, Adiyogi poured his profound knowledge into the legendary Saptarishis or "seven sages”. The sages carried this powerful yogic science to different parts of the world, including Asia, the Middle East, Northern Africa and South America. Interestingly, modern scholars have noted and marvelled at the close parallels found between ancient cultures across the globe. However, it was in India that the yogic system found its fullest expression. Agastya, the Saptarishi who travelled across the Indian subcontinent, crafted this culture around a core yogic way of life[2].

Yoga contains different asanas which speak to actual static stances. The utilization of posture assessment for yoga is trying as it includes complex setup of stances. Moreover, some best in class strategies neglect to perform well when the asana includes even body act or then again when both the legs cover one another. Subsequently, the need to build up a strong model which can help advocate self-taught yoga frameworks emerges.

Humans are prone to musculoskeletal disorders with aging and accidents. In order to prevent this some, form of physical exercise is needed. Yoga, which is a physical and spiritual exercise, has gained tremendous significance in the community of medical researchers. Yoga has the ability to completely cure diseases without any medicines and improve physical and mental health . A vast body of literature on the medical applications of yoga has been generated which includes positive body image intervention, cardiac rehabilitation, mental illness etc.. Yoga comprises of various asanas which represent physical static postures[1].

II. LITERATURE SURVEY

Efforts reported in this literature survey have focused on all the categories of Yoga pose detection and correction systems. Here we present a survey for Yoga pose detection and correction system.

Chen et al., 2018 [1] proposed a Yoga self-training system to assist in rectifying postures while performing Yoga using a Kinect depth camera for 12 different asanas. However, it is using manual feature extraction and making separate models for each asana. Chen HT, 2013 [7] also provided computer-assisted self-training system for posture rectification using Kinect. It has taken three postures in consideration, i.e. tree, warrior III, and downward facing dog. However, the overall accuracy is very low at only 82.84%.[1]

Y. Zhang et al., 2015 [4] address the localization problem by 1) using a search algorithm based on Bayesian optimization that sequentially proposes candidate regions for an object bounding box, and 2) training the CNN with a structured loss that explicitly penalizes the localization inaccuracy. They demonstrated that each of the proposed methods improves the detection performance over the baseline method on PASCAL VOC 2007 and 2012 datasets.[2]

Wu et al., 2010 [16] proposed an image and text-based expert system for Yoga. However, they have not analyzed the practitioner’s posture. Trejo EW and Yuan P, 2018 [17] proposed a Yoga detection system is proposed for six asanas using Kinect and Adaboost classification with 94.78%

accuracy score. However, they are using depth sensor-based camera which generally may not be available to the users.[3]

S. Gupta et al., 2014 [6] proposed a new geocentric embedding for depth images that encodes height above ground and angle with gravity for each pixel in addition to the horizontal disparity. They demonstrated that this geocentric embedding works better than using raw depth images for learning feature representations with convolutional neural networks. The final object detection system achieved an average precision of 37.3%, which is a 56% relative improvement over existing methods.[4]

B. Sapp et al., 2013 [10] proposed a multimodal and decomposable model for articulated human pose estimation in monocular images. Here they used a model of human pose that explicitly captures a variety of pose modes. Unlike other multimodal models, their approach includes both global and local pose cues and uses a convex objective and joint training for mode selection and pose estimation. They also employed a cascaded mode selection step which controls the trade-off between speed and accuracy, yielding a 5x speedup in inference and learning.[5]

Y. Tian et al., 2012 [12] proposed a new hierarchical spatial model that can capture an exponential number of poses with a compact mixture representation on each part. Using latent nodes, it can represent high-order spatial relationship among parts with exact inference. Different from recent hierarchical models that associate each latent node to a mixture of appearance templates (like HoG), they use the hierarchical structure as a pure spatial prior avoiding the large and often confounding appearance space.[6]

C. Lonescu et al., 2011 [15] presented an approach for automatic 3D human pose reconstruction from monocular images, based on a discriminative formulation with latent segmentation inputs. They advanced the field of structured prediction and human pose reconstruction on several fronts. By working with a pool of figure-ground segment hypotheses, the prediction problem is formulated in terms of combined learning and inference over segment hypotheses and 3D human articular configurations.[7]

S. Johnson et al., 2010 [20] investigated the task of 2D articulated human pose estimation in unconstrained still images. This is extremely challenging because of variation in pose, anatomy, clothing, and imaging conditions. They show that simple models of body part appearance and plausible configurations severely limit accuracy. They propose richer models of both appearance and pose, using state-of-the-art discriminative classifiers without introducing unacceptable computational expense.[8]

Shakhnarovich et al., 2003 [21], who how-ever use locality sensitive hashing. More recently, Gkioxariet al., 2013 [22] propose a semi-global classifier for part configuration. This formulation has shown very good results on real world data. However, it is based on linear classifiers with less expressive representation and is tested on arms only. Finally, the idea of pose regression has been employed by Ionescu et al., 2011 [23], however they reason about 3D pose.[9]

G. Mori et al., 2002 [25] proposed a system to estimate human body configurations using shape context matching. They have taken a single two-dimensional image containing a human body, locate the joint positions, and use these to estimate the body con-figuration and pose in three-dimensional space. The basic approach is to store a number of exemplar 2D views of the human body in a variety of different configurations and viewpoints with respect to the camera. On each of these stored views, the locations of the body joints (left elbow, right knee etc) are manually marked and labeled for future use. The test shape is then matched to each stored view, using the technique of shape context matching.[10]

III. EXISTING SYSTEM

At present, there are many yoga classes being conducted across many platforms. But in all these cases, there’s always a need for the mentor to be present at all times for guidance. Also we have to attend these classes at the mentioned time which may ruin our day’s schedule and we may end up missing yoga classes on some days. This restricts us from accessing these classes whenever we want to as the mentor cannot be always present to guide us. Also we can see that although there are many yoga classes, but the good ones always seem to be expensive. So, we would like to develop a yoga pose detection and correction system in real time which will be cost efficient and can be accessed by anyone at any time.

IV. PROPOSED SYSTEM

The proposed solution will develop a yoga pose correction and detection system using deep learning. We introduce a system using OpenPose followed by CNN models that allows us to get the keypoint detection of the user’s yoga pose within few seconds without retraining the model. The detected keypoints are then passed to the model which can easily predict the user’s yoga pose. Then a comparision is done between the keypoints of the user’s pose and that of the target pose. If a less similarity score is achieved, then verbal instructions are given to the user for pose correction. On top of that, after work-out reviews should be available as well so that the users can learn from their past mistakes and improve on their future workouts. Finally, in order to not impede the user’s movement and potentially increase the likelihood of injury, the solution does not require the practitioner to put on any additional equipment such as wearable sensors.

PROPOSED SYSTEM ARCHITECTURE

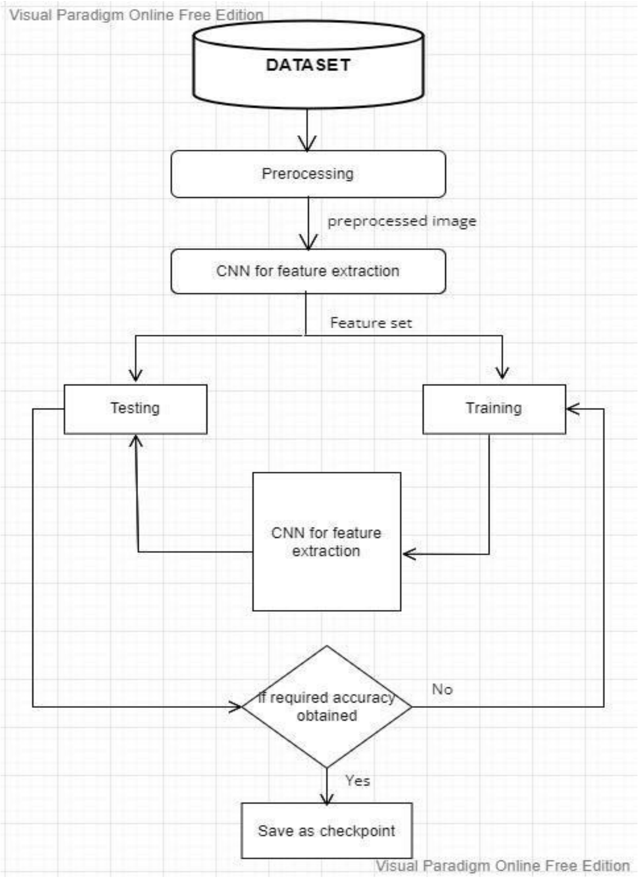
Visual Paradigm Online Free Edition
DATASET
Prerocessing
preprocessed image
CNN for feature extraction
Feature set
Testing
Training
CNN for feature extraction
If required accuracy obtained
No
Yes
Save as checkpoint
Visual Paradigm Online Free Edition

Fig 1 : Proposed System Architecture for finding the accuracy of the image processed using the CNN algorithm.

V. METHODOLOGY USED

The methodology will be prototype model of system development life cycle which is iterative andwill involve the stages of system analysis, system design, implementation/development, testing and maintenance. The proposed system can be implemented using SDLC life cycle. It is a framework defining tasks performed at each step in the software development process. It consists of a detailed plan describing how to develop, maintain and replace specific software. The life cycledefines a methodology for improving the quality of software and the overall development process.

A. Capture user movements

Video feed from the webcam using OpenCV is used to capture the user movements for theyoga pose detection and correction. OpenCV is a library of programming functions mainly aimedat real-time computer vision. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.

openPose:

Here we are using the OpenPose, the location of human body joints can be obtained from an RGB camera (Fig. 2). The keypoints obtained using OpenPose include ears, eyes, nose, neck, shoulders, elbows, wrists, knees, hips, and ankles as shown in Table 1. It can process inputs from a real-time camera, recorded video, static images, IP camera, etc., and present the results as 18 simple keypoints. This makes it suitable for a wide range of applications including surveillance, sports, activity detection, and Yoga pose recognition.

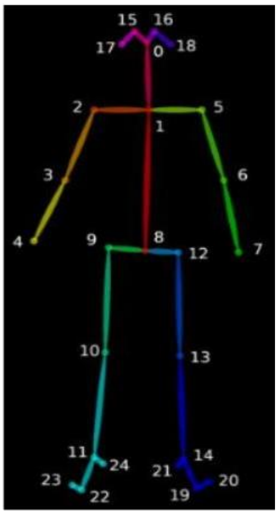

Fig 2: Keypoint detection using openPose

B. Normalising the Inputs

we are normalizing data because it improves the accuracy of the model very much. normalizing means to fix therange of the image. As, generally image has the range between0 to 255. As this range is quite large, we limit this range of theimage by division method to -1 to 1. As 0 means black and 255means white and colors basically represents an image. To limitthis range of (0,255) to (-1,1), we divide the original range i.e. 0 to 255 by 127 and then subtract it by 1.

C. Data Augmentation

We augment the images by performing scales, rotation, sheering or basically we can say we are pre-processing the images before it goes for training and testing.

D. Feature Extraction

In CNN networks images are taken as input and features are extracted by one layer and utilized by other layer. Extracted features are passed forward to classify images in CNN network. The next step after this is image classification. Here we extract features of images by:

1. Convolution
2. Non-Linearity
3. pooling

E. Image Classification using CNN

After the feature extraction using CNN we need to classify the images or videos (i.e., collection of frames). CNN uses various layers to classify the images. Here we classify images by using:

1. Flattering
2. Fully connected neural network
3. Softmax

F. Input to the Project

We are giving input to this yoga pose detection project as

- Image- We can give input as image i.e., the image of thepose that the user is doing
- Real time video- We can open webcam to capture real timevideo

G. Expected output

We are expecting the output of this yoga pose detection project is that we are getting a recognized image or video ofthe pose of the yoga of the user with the description of the name of that yoga pose and Accuracy.

H. Real-Time pose detection Results

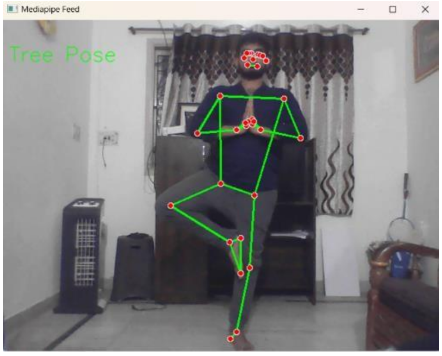

Fig 3 : Identifying as a Tree pose

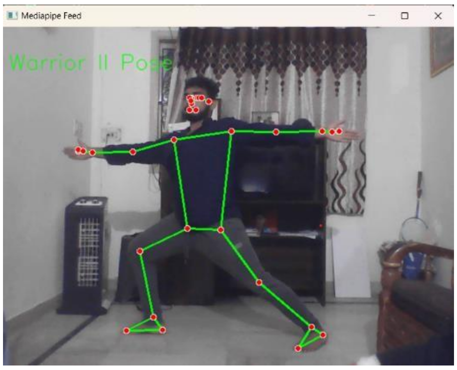

Fig 4 : Identifying as warrior II pose

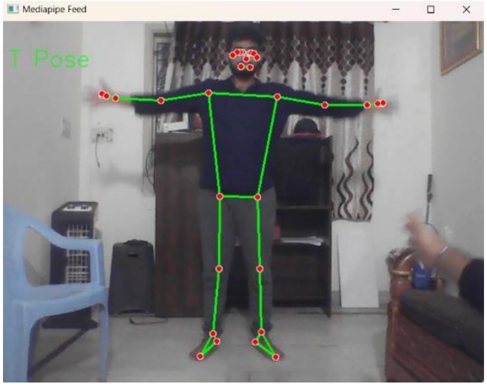

Fig 5 : Identifying as a T-pose

VI. IMPLEMENTATION

Convolutional Neural Network (CNN) :

Convolutional Neural Networks (CNNs) have been widely used as a key deep learning algorithm. CNNs are particularly effective for tasks involving image analysis and recognition due to their ability to automatically learn hierarchical features from input images.

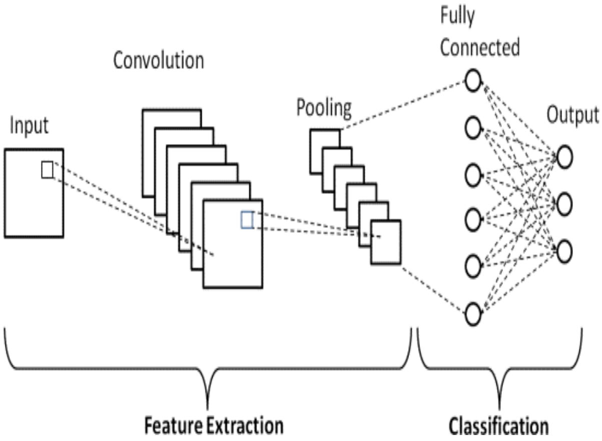

Fig 6 : Image classification process in CNN

1. Input Preprocessing:

The input images or frames, typically obtained from cameras or sensors, are preprocessed to enhance the relevant features for posture detection and correction.

2. Convolutional Layers:

The convolutional layers are the core building blocks of CNNs. They consist of a set of learnable filters (kernels) that convolve over the input images to extract local features.

3 Activation Functions:

Activation functions, such as ReLU (Rectified Linear Unit), are applied to introduce non-linearity into the network, allowing it to learn complex representations.

4 Pooling Layers:

Pooling layers are used to downsample the feature maps, reducing their spatial dimensions while preserving the most salient features.

5 Fully Connected Layers:

The output of the convolutional layers is flattened and connected to fully connected layers, which perform high-level feature learning and decision-making.
Fully connected layers are composed of densely connected nodes, where each node represents a feature or class label.

6 Training and Optimization:

The CNN model is trained using labeled data, where input images are associated with corresponding correct posture labels or corrective actions.
The model parameters (weights and biases) are optimized using backpropagation and gradient descent algorithms to minimize a loss function, such as categorical cross-entropy or mean squared error.

7 Posture Detection:

Once the CNN model is trained, it can be used for human posture detection by inputting new images or frames and obtaining predicted posture labels.
So, after the implementation process, the model accuracy is shown in following graph .

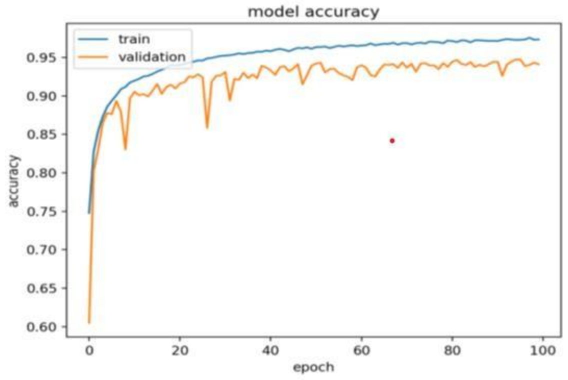

Fig 7 : OpenPose CNN Model training and validation accuracy

From the plot of accuracy we can see that the model is trained very well as the trend for accuracy on datasets is increasing as the epochs is increased and achieved 0.96% accuracy . We can also see that the model has not yet over-learned the training dataset, showing comparable skill on datasets.

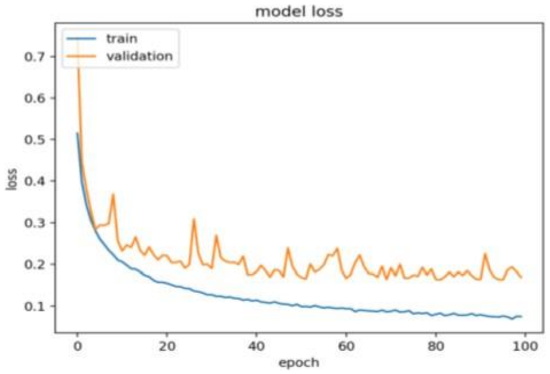

Fig 8 : OpenPose CNN Model training and validation loss

From the plot of loss, we can see that the model has comparable performance on both train and validation datasets (labeled test). If these parallel plots start to depart consistently, it might be a sign to stop training at an earlier epoch.

VII. CONCLUSION AND FUTURE WORK

We developed a project for the Yoga pose detection and correction system which will be able to operate without any sensors. The literature review helped us understand the drawbacks of the existing systems. As compared to other computer vision problems, human pose estimation is different as it has to localize and assemble human body parts on the basis of an already defined structure of the human body. Our aim is to present a solution that can help user practice yoga poses and correct their poses. Yoga self-instruction systems carry the potential to make yoga popular along with making sure it is performed in the right manner.

OpenPose is restricted by the way that 2D pictures can just view somebody from one point. Subsequently, our technique doesn't function admirably when parts of the body are obscured. In this way, it will be gainful to investigate location of postures during a 3D portrayal of the human body. Also our model presently centers on distinguishing the asana posture of a solitary individual. Nonetheless, OpenPose does accompany the abilities of distinguishing postures of different people in a picture. Thus, in future we might want to work on our model to distinguish different individuals in a single casing. This would be helpful for a gathering of individuals that are attempting to learn or rehearse yoga simultaneously.

REFERENCES

[1] Chen HT, He YZ, Hsu CC (2018) Computer-assisted yoga training system. Multimed Tools Appl 77:23969–23991. https://doi.org/10.1007/s11042-018-5721-2

[2] Y. Zhang, K. Sohn, R. Villegas, G. Pan, and H. Lee, “Improving object detection with deep convolutional networks via bayesian optimization and structured prediction,” inCVPR, 2015.

[3] Wu W, Yin W, Guo F (2010) Learning and self-instruction expert system for Yoga. In: Proceedings of 2010 2nd International Work Intelligent System Application: ISA, pp 2–5. https://doi.org/10.1109/iwisa.2010.5473592

[4] S. Gupta, R. Girshick , P. Arbel ́aez, and J. Malik, “Learning rich features from rgb-d images for object detection and segmentation,” inECCV,2014.

[5] B. Sapp and B. Taskar. Modec: Multimodal decomposable models for human pose estimation. InCVPR, 2013.

[6] Y. Tian, C. L. Zitnick, and S. G. Narasimhan. Exploring thespatial hierarchy of mixture models for human pose estimation. InECCV, 2012.

[7] C. Lonescu, F. Li, and C. Sminchisescu. Latent structured models for human pose estimation. InICCV, 2011.

[8] S. Johnson and M. Everingham. Clustered pose and non-linear appearance models for human pose estimation. InBMVC,2010.

[9] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter-sensitive hashing. InCVPR, 2003.

[10] G. Mori and J. Malik. Estimating human body configurations using shape context matching. InECCV, 2002.

[11] Toshev A, Szegedy C (2013) DeepPose: human pose estimation via deep neural networks. https://doi.org/10.1109/cvpr.2014.214

[12] Trejo EW, Yuan P (2018) Recognition of Yoga poses through an interactive system with kinect device. In: 2018 2nd international conference robotics and automation science: ICRAS, pp 12–17. https://doi.org/10.1109/icras.2018.8443267

[13] Mohanty A, Ahmed A, Goswami T et al (2017) Robust pose recognition using deep learning. In: Raman B, Kumar S, Roy PP, Sen D (eds) Advances in intelligent systems and computing. Springer, Singapore, pp 93–105. https://doi.org/10.1007/978-981-10-2107-7_9

[14] Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2016). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. ArXiv:1611.08050 [Cs]. Retrieved from http://arxiv.org/abs/1611.08050

[15] Wu W, Yin W, Guo F (2010) Learning and self-instruction expert system for Yoga. In: Proceedings of 2010 2nd International Work Intelligent System Application: ISA,pp 2–5. https://doi.org/10.1109/iwisa.2010.5473592