Yoga Pose Detection and Correction

S.C. Sethi[1], Sampada Bhosale[2], Prachi Dawkhar[3], Mitali Rote[4]

[1]*Assistant Professor, Department of Computer Engineering, NBNSTIC, Pune, India*
[2,3,4] *Student, Department of Computer Engineering, NBNSTIC, Pune, India*

---***---

Abstract - Yoga is a practice that connects body, breath, and mind. Yoga is extremely popular due to its numerous mental and physical health benefits. However, it is important to perform yoga poses correctly to maximize the benefits of yoga and prevent injury. To address this requirement, a yoga pose estimation and correction model was developed using real-time video input and pose detection and correction algorithms. Yoga Pose Detection and Correction is an innovative project that uses computer vision and artificial intelligence to improve yoga practice. The purpose of this research is to develop deep learning-based technology to estimate the correct posture of the practitioner and make the task easier. To implement this strategy, the study used his four different deep learning architectures: MediaPipe, OpenPose, PoseNet, and EpipolarPose. The main components of the project include advanced video processing algorithms and pose recognition models. CNN is a network architecture for deep learning algorithms and is used for image recognition. Comparing the estimation accuracy of all architectures in the study, the MediaPipe architecture provides the highest estimation accuracy.

Key Words: Artificial intelligence, deep learning, machine learning techniques, pose estimation techniques, skeleton and yoga.

1. INTRODUCTION

Yoga pose detection and correction utilizes a real-time video input and the fastest algorithm for the detection and correction of poses. The recent development of visual pose estimation algorithms for both 2D and 3D has created new opportunities to use smartphones and webcams in accurately determining human body positions. The application utilizes an artificial intelligence-based system that not only just detects yoga poses but also provides personalized feedback to individuals to help them to improve their poses. It analyzes a large dataset of different yoga poses to outline the skeleton of the human body during practice. Corrections are provided for each step to ensure accuracy in posture. Yoga pose estimation is a complex problem used in computer vision-based approaches. This involves identifying and tracking key body points such as joints and limbs with real-time video input.

2. Body of Paper

2.1 Problem Statement:

As yoga becomes more popular around the world, there is an urgent need for intelligent systems that can accurately identify and modify yoga poses during practice sessions. A variety of yoga platforms and applications are available, but many are unable to provide personalized feedback on a practitioner's alignment and form in real time. The potential health benefits of yoga are undermined by the lack of reliable systems for detecting and correcting yoga postures. Because incorrect posture can lead to injury and reduce the overall effectiveness of the exercise.

2.2 Methodology:

Our method aims to automatically recognize customers' yoga poses from real-time and recorded videos. The implementation is divided into four techniques. First, a process is performed to collect the necessary data. This is either a continuous cycle that runs in parallel with identification or recently recorded video. Next, use OpenPose to find joint positions using part trust maps and part affinity fields. This process involves identifying and assigning corresponding body regions in a two-step manner and then analyzing the resulting data. The identified keypoints are passed to the model, the CNN finds examples, and after some time the LSTM checks for changes. Finally, we consider models and preparation strategies for frame-by-frame expectations and measurement approaches for 45 yield edges.

Pose Extraction:

This is the first step in the pipeline, and it uses the OpenPose library. This progress happens offline because recording occurs, whereas real-time forecasting involves continuous video recording. The proposed model uses web-based inputs from cameras to flexibly extract keypoints. OpenPose is a library that detects key points of multiple people, such as the human body, hands, and faces. Location of 18 important points such as ears, eyes, nose, neck, shoulders, hips, knees, lower legs, elbows, wrists, etc. The yield of each frame of the video is determined in his JSON format, which contains the area of each body part of each person detected in the image. For optimal performance, we extracted poses using OpenPose's default settings. The framework ran at around 3 FPS.

Model:

The deep learning model used here is a combination of CNN and LSTM. CNNs are used for image recognition tasks, while LSTMs are often used for time series analysis. Here, we used a time-cyclic CNN layer to extract highlights from the 2D directions of the keypoints obtained in the previous step. The LSTM layer analyzes feature changes across frames and assigns a probability to each yoga pose using a softmax layer. We analyze the effect of surveying on 45 edges by applying a threshold to distinguish the contours of non-yoga activities. This model was refined using Python's Keras Sequential API to improve performance. The input data consists of 45 continuous edges and 18 keypoints, with each keypoint having an X and Y coordinate. A time-distributed CNN layer with 16 filters of size 3x3 is applied to extract features from the keypoints of each edge. To speed up the combination, cluster normalization is applied to the CNN yield. The video input is analyzed sequentially, treating the entire yoga session as a sequence from initial development to implementation and execution.

Training:

Our task is to gradually recognize the client's asanas with appropriate precision. Before applying CNN and LSTM models to Asana predictions, we first use OpenPose to remove keypoint highlights and record joint region estimates in a JSON file. Combining both results in the optimal placement of highlights filtered by CNN and the condition of long-range information established by LSTM. Keras is used to aggregate models in the Theano backend.

2.3 System Design:

Input Data: The system receives input from a camera or video stream to capture a yoga practitioner performing various poses.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Computer Vision: OpenCV (an opensource computer vision library) is used for image and video processing. It provides functionality for tasks such as reading video images, preprocessing images, and displaying processed output.

Pose Estimation: Mediapipe is a library developed by Google that provides a comprehensive set of tools for building various recognition pipelines, including pose estimation. The system uses the Mediapipe Pose module to detect and estimate key points (landmarks) on the human body, such as joints and other body parts.

Landmark Detection: The pose estimation model identifies specific landmarks on the body, including major joints such as shoulders, elbows, wrists, hips, knees, and ankles.

Yoga Pose Recognition: Based on the detected landmarks, the system uses a set of predefined rules or a machine learning model to recognize the yoga pose being performed. This may include comparing the current pose to a database of reference poses.

Pose Correction: Provides feedback or corrections when the system detects that the user's pose deviates from correct form. This feedback can be visualized in real time.

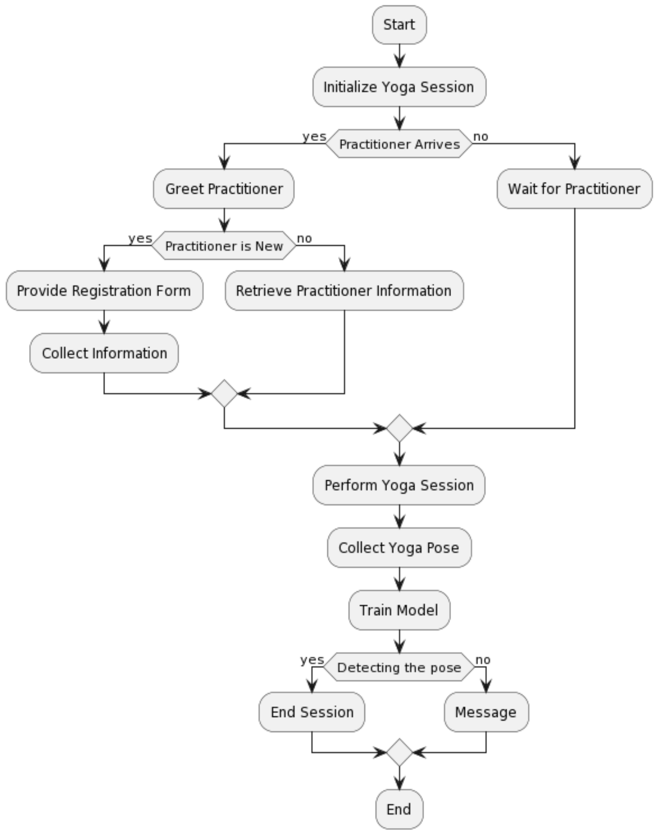

Fig 1. System Architecture

Component Details:

1. **OpenPose:**

OpenPose is a continuous multi-individual keypoint position that has revolutionized the field of posture assessment. It was created by the Perceptual Registration Institute at Carnegie Mellon University (CMU).
Use CNN-based techniques to identify points on the human body's hands, feet, and face from a single image. OpenPose uses an RGB camera to identify joints in the human body. The eyes, ears, neck, nose, elbows, shoulders, knees, wrists, lower legs, and hips are among the OpenPose keypoints. It provides eighteen fundamental keypoints that summarize the results obtained from processing live camera input, pre-recorded recordings, or still images. The work presented in makes use of CNN to characterize yoga presents after using

OpenPose to provide introductory keypoint recognizable proof.

2.**PoseNet:**

PoseNet is a deep learning framework similar to OpenPose that is used to identify human postures from images or videos by identifying the different parts of the body that each joint makes. The "Part ID" score, which is a positive number with a range of 0.0 to 1.0 and 1.0 being the best is used to list these joint areas or keypoints. PoseNet model execution varies depending on the device and yield step. Because the PoseNet model is size invariant, it can predict the current situation in the original image dimensions, regardless of whether the image has been downscaled or not.

2.4 Applications:

1. **Virtual Yoga instructor:** Creating virtual yoga instructors that can guide users through poses with real-time feedback. Offering personalized tips and corrections based on performance.

2. **Programs for Corporate Wellness:** Integrating yoga pose detection into workplace wellness initiatives to support workers' well-being and lower stress levels offering customised yoga programmes to staff members according to their individual requirements.

3. **Monitoring Well-Being and Health:** Yoga poses observed and evaluated for patients receiving physical rehabilitation giving medical professionals knowledge about a patient's general physical health, balance, and flexibility.

4. **Education and Training**: Yoga Education provide a pose analysis tool to help yoga instructors in their training programs. Make use of the system to instruct others by modeling appropriate yoga poses and techniques.

3. Literature survey

AI techniques may rely on heuristic human element extraction for routine tasks like finding social exercise locations. Human zone awareness places restrictions on it. A few strategies, such as deep learning approaches, have been chosen by the creators to discuss this risk. Accordingly, during the preparation phase, these tactics could extract specific highlights from raw sensor data, and subsequently low-level transient attributes with significant level unique requests would be introduced.

1) Paper name: Real-time Yoga recognition using deep learning
Author name: Santosh Kumar Yadav, Amitojdeep Singh, Abhishek Gupta, Jagdish Lal Raheja
Description: In this paper, a hybrid deep learning model combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) is presented as a method for identifying and categorizing yoga poses. With single frame testing, the proposed system achieves 99.04% test accuracy, and with polling of predictions on 45 frames of the videos, it achieves 99.38% accuracy.

2)Paper name: PifPaf: Composite Fields for Human Pose Estimation
Author name: Sven Kreiss, Lorenzo Bertoni, Alexandre Alahi
Description The study presents a novel technique for multiperson 2D human pose estimation called PifPaf, which is especially well suited for urban mobility applications like autonomous vehicles and delivery robots. The technique localizes body parts using a Part Intensity Field (PIF) and a Part using the Association

Field (PAF), body parts can be linked to create complete human stances.

3)Paper name: ExNET: Deep Neural Network
for Exercise
Author name: Sadeka Haque, AKM Shahariar Azad Rabby, Monira Akter Laboni, Nafis Neehal, and Syed Akhter Hossain
Description: The application of convolutional neural networks (CNN) for human pose detection, specifically for exercise poses, is the main topic of the research paper. In addition to highlighting the difficulties in human pose analysis, the paper emphasizes the importance of human pose classification in computer vision applications such as augmented reality, fitness, and animation.

4)Paper name: XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera
Author name: Dushyant Mehta, Oleksandr Sotnychenko, Franziska Mueller, Weipeng Xu, Mohamed Elgharib, Pascal Fua, Hans-Peter Seidel, Helge Rhodin, Gerard Ponsmoll, Christian Theobalt
Description The research presents a real-time approach for multi-person 3D motion capture using a single RGB camera. The system operates successfully in generic scenes, handling occlusions and interactions in general scene settings, and localizing subjects relative to the camera. The paper introduces a three-stage approach for capturing 3D motion in real-time, involving a convolutional neural network (CNN) for 2D and 3D pose estimation, a fully-connected neural network for further 3D pose estimation, and a space-time skeletal model fitting for temporal coherence.

5)Paper name: "Yoga Pose Assessment Method Using Pose Detection for Self-Learning"
Author name: M. C. Thar, K. Z. N.
Description: This paper suggests a method for evaluating yoga poses in order to use pose detection to support the process of self-mastery in yoga. This study suggested a Yoga Pose Training System with Performance Evaluation to help with self-yoga mastery. This article provides information on how to find yoga poses and how to use pose discovery to support introspective yoga practice.

CONCLUSION AND FUTURE SCOPE

The approach presented in this study is based on deep learning to detect and correct yoga postures and recommend the user to improve their posture by indicating where the yoga postures are incorrect. The proposed system aims to Our method aims to automatically recognize customers' yoga poses from real-time videos. The combination of CNN and LSTM model applied to OpenPose data appears to be extremely effective in organizing yoga poses flawlessly. Additionally, a basic CNN and SVM perform far better than we would have hoped. The application of SVM shows that machine learning calculations can also be used for problems related to movement recognition or assessment. Additionally, compared to a neural network, SVM is much lighter, less unpredictable, and requires less resources.

Multi-individual posture assessment is a very distinct topic in and of itself, with a great deal of potential for further investigation.

More asanas and a larger dataset with both images and recordings can be included in subsequent work. Furthermore, for ongoing expectations and self-preparation, the framework can be used on a handy device. This piece of work serves as a stand-in for an acknowledgment of action.

ACKNOWLEDGEMENT

We are grateful to NBN Sinhgad Technical Institute Campus, Pune, for constant support and encouragement in working on our paper, and we are also thankful to Prof. S.B. Bendale (HOD Comp. dept) and S.C. Sethi (Guide) for their helpful guidance and support in improving our paper.

REFERENCES

[1]. L. Sigal. “Human pose estimation”, Ency. of Comput. Vision, Springer 2011.
[2]. S. Yadav, A. Singh, A. Gupta, and J. Raheja, “Real-time yoga recognition using deep learning”, Neural Comput. and Appl., May 2019.Online]. Available: https://doi.org/10.1007/s00521-019-04232-7
[3]. U. Rafi, B. Leibe, J.Gall, and I. Kostrikov, “An efficient convolutional network for human pose estimation”, British Mach. Vision Conf., 2016.
[4]. S. Haque, A. Rabby, M. Laboni, N. Neehal, and S. Hossain, “ExNET: deep neural network for exercise pose detection”, Recent Trends in Image Process. and Pattern Recog., 2019.
[5]. M. Islam, H. Mahmud, F. Ashraf, I. Hossain and M. Hasan, "Yoga posture recognition by detecting human joint points in real time using microsoft kinect", IEEE Region 10 Humanit. Tech. Conf., pp. 668-67, 2017.
[6]. A. Agarwal and B. Triggs, “3D human pose from silhouettes by relevance vector regression”, Intl Conf. on Computer Vision & Pattern Recogn.pp.882–888, 2004. Available: https://arxiv.org/pdf/1907.05272.pdf.