Body Posture In Self Learning Activities: A Comprehensive Review & Analysis

Prof. Roshni Narkhede [1], Saurabh Sonalkar[2], Dhiraj Yadav[3]
Computer Engineering Department[1,2,3]
Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra[1,2,3]

Abstract —This study presents an innovative approach for accurately recognizing diverse Yoga poses using deep learning algorithms. Our proposed system introduces a method for Yoga pose assessment leveraging pose detection to facilitate self-learning of Yoga. Utilizing multi-parts detection solely with a PC camera, the system effectively identifies Yoga poses. Furthermore, we introduce an enhanced algorithm for scoring applicable to all poses. Evaluation of our application encompasses various Yoga poses across different scenarios, demonstrating its robustness. We propose a hybrid deep learning model incorporating linear regression for real-time Yoga recognition in videos. This model utilizes linear regression to extract features from key-points in each frame, obtained through Open-Pos.

Keywords— Yoga Pose Recognition, Deep Learning Algorithms, Linear Regression, Real-time Video Analysis, Pose Detection System

INTRODUCTION

Human pose estimation poses a significant challenge within the realm of computer vision. It involves pinpointing the location of human joints in images or videos to create a skeletal representation. Automatically detecting a person's pose is complex due to various factors such as image scale, resolution, lighting conditions, background clutter, clothing variations, and interaction with surroundings.

Pose estimation finds applications in areas likeexercise and fitness, which are increasingly pertinent given today's busy lifestyles.[7] People face numerous health issues due to hectic schedules, highlighting the importance of adopting a healthy lifestyle encompassing nutritious diet, physical activities, weight management, and stress reduction. Engaging in exercises, including aerobics, strength training, balance exercises, cardio, and yoga, is crucial for maintaining health.

While yoga offers spiritual, physical, and mental benefits, practicing it incorrectly can lead to adverse outcomes.[5] Many individuals lack access to instructors, underscoring the potential for artificial intelligence-based applications to identify yoga poses and offer personalized feedback.

This project aims to explore various approaches foryoga pose classification, addressing the need for accurate posture assessment and guidance in self-directed yoga practice.

Yoga, as an ancient practice, holds profound significance due to its multifaceted benefits encompassing spiritual, physical, and mental well-being.[2] Its holistic approach to health resonates with individuals seeking a between mind and body. Through a combination of postures, breathingexercises, and meditation, yoga offers a pathway to inner peace and balance amidst the chaos of modern life. However, mastering the intricacies of yoga poses requires patience, dedication, and guidance to ensure alignment and avoid potential injuries.

In recent years, technological advancements have revolutionized the way we approach health and wellness. The emergence of wearable devices, mobile applications, and virtual platforms has democratized access to fitness resources, empowering individuals to take charge of their well-being.[3] Integrating artificial intelligence into fitness solutions presents an exciting opportunity to enhance the efficacy and accessibility of yoga practice. By leveraging machine learning algorithms and computer vision techniques, AI-powered applications can analyze user movements, provide real-time feedback, and tailor personalized yoga routines to suit individual needs and preferences.

Moreover, the fusion of yoga and technology extends beyond individual practice to community engagement and global connectivity.[8] Online platforms and social media channels serve as virtual hubs for yoga enthusiasts worldwide to share experiences, exchange knowledge, and participate in collaborative initiatives. Through live streaming sessions, virtual workshops, and interactive forums, practitioners can connect with like-minded individuals,learn from experienced instructors, and cultivate a sense of belonging within a supportive community.[1] This digital ecosystem fosters inclusivity, diversity, and innovation, paving the way for a more interconnected and empowered global yoga community.

PROBLEM DEFINITION

Yoga pose estimation is a significant challenge in computer vision, requiring the automatic detection of a person's pose in images.[4] This task is complex due to factors like image scale, resolution, lighting, background clutter, clothing variations, and interactions with surroundings.

The diversity of yoga asanas compounds the difficulty of creating a successful pose estimation model. With numerous pose variations and transitions, ensuring accuracy across all poses presents a considerable challenge.

Addressing these challenges demands innovative approaches in computer vision and machine learning,

along with comprehensive datasets and collaborative efforts among researchers, yoga practitioners, and technology developers[8].

I. OBJECTIVE

The project aims to utilize image processing techniques, particularly Convolutional Neural Networks (CNN), for the identification of human actions. This involves analyzing visual data to recognize various activities performed by individuals.[1] CNNs are selected due to their capability in capturing spatial hierarchies within images, making them well-suited for tasks like action recognition.

Specifically focusing on Yoga, the project endeavors to utilize image processing for the detection and refinement of Yoga poses. Its objective extends beyondmere recognition to providing corrective guidance, thereby aligning with the broader aim of enhancing Yoga practice through technological means.[1]

An integral component of the project entails the examination and extraction of optical flow from image data. Optical flow, which depicts motion patterns in a sequence of images, is under scrutiny.[2] The project aims to develop a CNN model proficient in handling optical flow data effectively, thus contributing to a deeper understanding of pose detection, especially in dynamic contexts.

The project underscores the importance of training a deep learning model to accurately classify various Yoga poses. This process involves instructing the model to effectively recognize and categorize poses. Emphasis on the learning process within the deep learning model signifies a dynamic adaptation, ensuring proficiency inhandling a wide array of Yoga poses.

Ultimately, the project's overarching goal is to train a customized deep learning model tailored specifically for assessing the accuracy of a person's Yoga pose. This requires integrating knowledge gained from various project objectives to create a personalized model capable of identifying poses and evaluating their correctness.[6] Such a model aligns perfectly with the project's objective of offering meaningful feedback on Yoga practice.

II. SYSTEM ARCHITECTURE

The system architecture is meticulously crafted to facilitate the precise recognition of Yoga poses through a multi-stage process. It commences with the Video Capture Module, which retrieves real-time video streams from a PC camera, initiating the processing pipeline for subsequent analysis.

Central to the architecture is the Pose Detection Module, employing sophisticated algorithms to identifykey body points or joints in each frame of the video. This pivotal step relies on pose detection techniques, which may involve established models like OpenPose or PoseNet, or a custom-trained deep learning model tailored to Yoga poses.

Following successful pose detection, the architecture directs its attention to the Feature Extraction Module. Here, Linear Regression is employed to extract meaningful features from the identified key points, forming a parameter set that characterizes each Yoga pose and contributing to the subsequent recognitionprocess.

Given the real-time nature of video streams, the architecture integrates a Real-time Recognition Module.This component ensures swift processing of video frames, enabling instantaneous feedback on recognized Yoga poses, vital for a seamless and interactive user experience.

To validate the system's effectiveness, mechanisms for Evaluation and Robustness Testing are included. The system undergoes testing with diverse Yoga poses under varying scenes, evaluating metrics such as accuracy, precision, recall, or F1 score to assess its robustness and performance.

Integrated seamlessly into the architecture is Open-Pos, contributing to the pose estimation process and enhancing the overall system's accuracy and effectiveness, although specific details are not explicitlyoutlined in the abstract.

Concluding the architecture is the Output Module, tasked with presenting the final recognized Yoga poses and their respective scores to the end user. This output can take the form of a graphical user interface or any other communication means, facilitating user interaction and feedback.

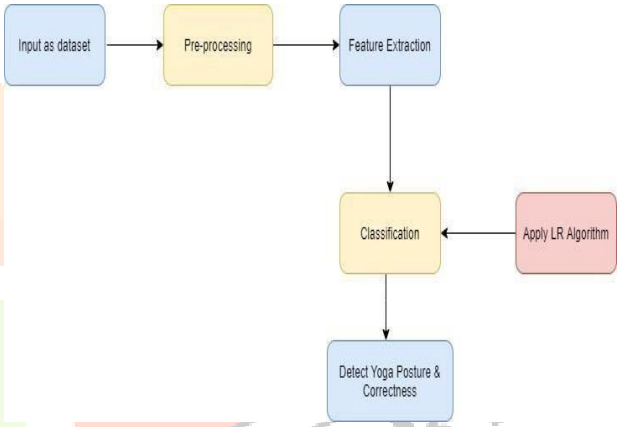

Figure 1: System Model

Data Collection- Gathering information on Yoga posesusing a PC camera, focusing on diverse poses andscenes to ensure comprehensive training data for the model.

Feature Extraction- Employing an improved algorithm to calculate scores based on key body points, enhancing the ability to identify and characterize various Yoga poses.

Training Models- Developing a hybrid deep learning model using Linear Regression to recognize Yoga poses in real-time videos, with a focus on efficient processing and accurate classification.

Similarity Learning- Integrating Open-Pos for pose estimation, contributing to the accuracy and effectiveness of the system in recognizing and assessingdifferent Yoga poses
.

Ranking System- Implementing a scoring mechanism to evaluate and rank the correctness of detected Yoga poses, aiming to provide users with meaningful feedback on their

III. LITERATURE SURVEY

The exploration of yoga pose detection in computer vision and human-computer interaction has garnered significant interest. This examination delves into various

methodologies, techniques, and advancements aimed at accurately detecting and categorizing yoga poses from images or videos. Understanding the latest developments in this field is crucial for the development of robust systems that can assist practitioners in refining their yoga practice, facilitating remote instruction, and improving healthcare operations.[3] The first paper, titled "An Overview of the Fashionability of Yoga," highlights theincreasing popularity of yoga and its associated physical, mental, and spiritual benefits. However, practicing yoga without proper guidance can lead to health issues such as strokes or nerve damage.

Therefore, it is essential to ensure correct posture during yoga sessions. The proposed system aims to identify and visually guide practitioners in real-time, using a vision-based approach.[4] The Continuous Yoga Instructor system captures practitioner movements using a mobile camera, streaming at 1280 x720 resolution and 30 frames per second to the detection system. The second paper discusses the integration of yoga into people's lives worldwide and the need for scientific analysis of yoga postures. Recognizing yoga postures in real-time poses a significant challenge due to limited datasets and computational constraints.[5] To address this, a large dataset comprising 5500 images of ten different yoga poses was created.

A pose estimation algorithm generates a skeletal representation of the human body, enabling joint angle extraction for use in machine learning models.[1] Training on 80% of the dataset and testing on the remaining 20% achieved an accuracy of 99.04% using a Random Forest Classifier.

.

IV. METHODOLOGIES

The Yoga Posture Detection and Correction System encompasses four primary components:

1. Keypoints Detection using OpenPose.
2. Keypoints Detection using Mask RCNN.
3. Higher Probability Prediction & Comparison.
4. Android Trainer Application.

The system workflow begins with capturing and streaming the user's movements in real-time to a media streaming server. Subsequently, the system employs a pose estimation library such as OpenPose or MaskRCNN to detect the user's joints or keypoints. These keypoints are then utilized in the yoga posedetection module to predict the user's pose before reaching the final phase of the asana. The results and accuracy values are transmitted to the android application through a data channel.

A. Dataset: A publicly available dataset for yoga postures comprises videos of six yoga poses performed by fifteen individuals. The videos, recorded at 1366 x 768 resolution and 30 frames per second using a standard web camera, depict common householdlighting settings. With a total of eighty-eight videos, each averaging 45 seconds, the dataset's duration amounts to approximately 1 hour and 6 minutes, totaling around 118,950 frames.

B. Keypoints Detection using OpenPose: The incoming video stream is processed by OpenPose, which detects 25 keypoints in the body. After feature extraction and generation of part confidence maps and part affinity fields, a greedy algorithm generates the output in JSON format containing body part locations for each frame. This output is then forwarded to the higher probability prediction and comparison component.

C. Keypoints Detection using Mask RCNN: Mask RCNN is utilized for human detection and keypoints detection. The input video stream undergoes feature extraction and region proposal generation, followed by RoI pooling and fully connected network application. Astacked hourglass network with spatial transformer network refines the generated pose proposals,producing output results in JSON format.

D. Higher Probability Prediction & Comparison: This component aims to predict the user's pose using the obtained "Joints JSON" data and provide feedback on the user's proximity to the original asana. Achieving minimal delay is crucial for a seamless user experience.

V. FUTURE SCOPE

The increasing recognition of Yoga's benefits has led to a burgeoning field in India and abroad, offering diverse career prospects for trained individuals. Beyond traditional settings like yoga studios, opportunities are expanding into special needs centers and private gyms, broadening the scope for Yoga practitioners to work with varied clientele.

Moreover, the fusion of Yoga with technology presents exciting prospects for the future. The development of Yoga pose recognition applications, such as fitness or yoga trackers, holds promise for personalized fitness monitoring and effective Yoga practice. For instance, applications tracking squats and push-ups exemplify how technology can automatically monitor user statistics, offering valuable insights into exerciseperformance and promoting data-driven fitness approaches.

This integration of Yoga and technology not only enhances personal well-being but also drives innovation in application development. As demand for holistic health solutions grows, there is potential for advanced applications beyond basic tracking. Future developments could include interactive platforms for virtual yoga sessions, AI-driven personalized workout plans, or gamified experiences to make fitness and yoga more engaging, appealing to a wider audience and amplifying Yoga's impact in the digital era.

Furthermore, beyond personal fitness, Yoga pose recognition technology can extend to healthcare and rehabilitation domains. Automatic assessment of exercise quality opens avenues for creating tools to aid healthcare

professionals in monitoring and guiding patients through rehabilitative yoga practices. This integration not only expands career opportunities for Yoga professionals but also promotes the incorporation of Yoga into mainstream healthcare practices, contributing to overall health and wellness.

VI. CONCLUSION

In summary, our methodology showcases a well-considered fusion of advanced neural network layers, notably integrating the time-distributed Convolutional Neural Network (CNN) layer to discern patterns among key points within individual frames. Concurrently, the strategic incorporation of the Long Short-Term Memory (LSTM) layer aims to capturepatterns observed across recent frames, thereby enhancing the system's comprehension of sequential movements inherent in Yoga poses.

The inclusion of LSTM in our model serves a dual role – facilitating pattern memorization across frames while mitigating errors arising from false key point detections. Additionally, the integration of pollingfor denoising further bolsters the system's resilience. By incorporating these components, our model exhibitsproficiency in handling the sequential aspect of Yoga images, resulting in more precise and dependable pose detection. The emphasis on sequential frames acknowledges the dynamic and continuous nature of Yoga movements, ensuring adaptability to the fluidity of Yoga practice.

Overall, the synergy between the time- distributed CNN layer, LSTM, and denoising through polling represents a holistic approach to optimizing system performance. This strategic amalgamation not only elevates accuracy but also fortifies the system against disruptions caused by false detections. As we explore future applications of this technology, the robustness achieved through these methodologies positions our system as a promising tool not only for Yoga pose detection but also for broader applications inmovement analysis, fitness tracking, and rehabilitation exercises.

VII. REFERENCES

1) V. Akuhota and S. F. Nedler authored a paper titled "Core Strengthening," published by the American Academy of Physical Medicine and Rehabilitation in 2004.

2) R. Szeliski authored "Computer Vision: Algorithmsand Applications," published by Springer in 2010.

3) G. Bradski and A. Kaehler authored "Learning OpenCV," published by O'Reilly in 2008.

4) Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh authored "Realtime Multi-Person 2D Pose Estimation usingPart Affinity Fields," published by The Robotics Institute, Carnegie Mellon University, in 2017.

5) P. Ganesh wrote an article titled "Human Pose Estimation: Simplified," published on Towards Data Science on March 26, 2019, accessible at https://towardsdatascience.com/humanpose-estimation-simplified6cfd88542ab3, accessed onApril 3, 2020.

6) T. Amert, N. Otterness, M. Yang, J. H. Anderson, and D. Smith authored "GPU scheduling on the NVIDIA TX2: Hidden details revealed," presented at the IEEE Real-Time Systems Symposium (RTSS) in December 2017.

7) B. Schoettle authored a report titled "Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles," published by the University of Michigan, Ann Arbor, MI USA

8) D. Feng et al., "Deep multi-model object detection and semantic segmentation for autonomous driving : Datasets, methods, and challenges," IEEE Trans. Tntell. Transp. Syst., early access, Feb. 17, 2020