

Pose Estimate Based Yoga Instructor

Aarti Bakshi¹, Danish Sheikh², Yasmin Ansari^{3*}, Chetan Sharma⁴, Harishchandra Naik⁵

^{1,2,3,4,5}Department of Electronics and Telecommunication Engineering, K. C. College of Engineering Management Studies and Research, Mumbai, India

Abstract: Wanting a private trainer to assist track our fitness goals, we have a tendency to patterned we have a tendency to might build our own. The goal was to make associate degree application that might track however we have a tendency to were physical exercise and commenced with Yoga as a straightforward context. we have a tendency to dubbed our 1st iteration of this application as Pose Estimation based yoga trainer to evaluate our construct, we start by gathering pictures of Yoga poses with a picture explore for terms like: 'yoga tree pose', 'yoga triangle pose', etc. we have a tendency to selected yoga as a result of the movements square measure comparatively static compared to different athletic maneuvers, this makes the constraints on frame rate of abstract thought less tightened. we will quickly separate extraneous photos and maybe refine our queries to make a corpus of one or two thousand yoga create pictures. Pose estimation gets USA half means. to comprehend Pose Estimation based yoga trainer, we'd like to feature one thing new. we'd like an operate that takes USA from cause estimates to yoga position categories. With up to fourteen body key points, every of our couple thousand pictures may be drawn as a vector in a very 28-dimensional real linear area. By convention, we'll take the x and y indices of the mode for every key purpose slice of our cause estimation model belief map. In different words, the cause estimation model can output a tensor formed like (1, 96, 96, 14) wherever every slice on the ultimate axis corresponds to a 96x96 belief map for the situation of a selected body key purpose. Taking the max of every slice, we discover the foremost doubtless index wherever that slice's key point is positioned relative to the framing of the input im-age. Our 1st plan here was to concatenate the create vectors from a pair of or three ordered time steps and take a look at to coach the tree to acknowledge a motion. to stay things easy, we have a tendency to begin by framing a need to differentiate be- tween standing, squatting, and forward bends (dead-lift). These classes were chosen to check each static and dynamic maneuvers. Squats and Dead-lifts continue to exist similar planes-of-motion and area unit leg-dominant moves though' activating opposing muscle teams.

Keywords: Estimation, Yoga tree pose, Yoga triangle pose.

1. Introduction

Yoga has Associate in Nursing ancient tradition of physical and mental coaching to enhance well-being. fashionable yoga, that emphasizes physical learning and mental strength through coaching physical postures, has been growing in quality over the last couple decades. However, the time and price of attending to a yoga studio is preventative. Still others would like to follow outside of the cluster setting. Here, we tend to explore the event of Yoga-AI.

With cause recognition, we tend to implement a wise assistant to supply corrective recommendation to guide practitioners. Our setup isn't restricted to Yoga flows. we are going to conjointly explore analysis and feedback for strength coaching movements. Advancements in edge devices specialized for machine learning training/inference on device and varied machine learning libraries rising abstract thought time on create estimation models gave approach for enhancements in Federal Protective Service for classifying poses and movement. We abstracted the technique of exploitation create estimation abstract thought out-put as input for associate LSTM classifier into a toolkit referred to as Action-AI. It's conspicuously featured in GitHub's act recognition topic and received a prize in NVIDIA's AI at the sting Challenge! Action-AI generalizes the approach of Yoga-AI associated connected comes framing a marsh plant pipeline by introducing trackers and multi-person create estimation. By baking create estimation into the pipeline be- cause the primary abstract thought engine, the developer will concentrate on coaching easy image classification models supported low dimensional options or little, localized image crops. Since in style marsh plant frameworks usually solely support the foremost common laptop vision tasks like object detection or image classification/segmentation, we wanted to implement our own. Many marsh plant frameworks use GStreamer to amass and method video. For our video process demo, Open-CV suffices. For create estimation we tend to use Open-pose enforced with in style deep learning frameworks like TensorFlow and PyTorch. Accurately recognizing some activities needs higher resolution in time with higher frame rates, thus we tend to use TensorRt converters for optimized abstract thought on edge-AI prototyping de- vices just like the Jetson Nano. The main programming abstraction of Action-AI may be a traceable person category, the same as this py images each traceable object. This object features a technique to enqueue the configuration of N (14 or 18) new key points as a length 2N num-py array into a circular buffer. For procedure potency, we tend to like smaller buffers, however we tend to balance this need with one to produce enough info as input for secondary models. This object additionally encapsulates ids, bounding boxes, or the results of running extra abstract thought.

*Corresponding author: ansariy183@gmail.com

2. Literature Review

There are ways or methodologies for Pose Estimation. We studied some of them. They are as follows,

We proposed a method for human pose estimation based on Deep Neural Networks (DNNs). The pose estimations formulated as a DNN-based regression problem towards body joints. We present a cascade of such DNN regressions which results in high precision pose estimates. This approach has advantage of reasoning about pose in a holistic fashion and has a simple and yet powerful formulation which capitalizes on recent advances in Deep Learning.

[1] The capacity to estimate the head pose of another person is a common human ability that presents a unique challenge for computer vision systems. In this paper, we are discussing the difficulties in head pose estimation and we present an organized survey describing the evolution of the field. Our discussion mainly focuses on the advantages and disadvantages of each approach and spans 90 of the most innovative and characteristic of papers that have been published on this topic. We compare these systems by focusing on their ability to estimate coarse and fine head pose, highlighting approaches that are very well suited for unconstrained environments

[2] H150 pose estimation is important research topic in the field of computer vision as well as artificial intelligence. This paper focuses on the state-of-art progress of 2-D human pose estimation methods based on the deep learning. According to the neural network structure, these methods are classified as single CNN method, Multi-stage CNN method, Multi-branch CNN method, Recurrent Neural Network (RNN) method and Generative Adversarial Networks (GAN) method. We summarize and analyze their attributes and performance. The future development direction is also prospected.

[3] Human pose estimation has always been a challenging problem that holds great attention, it has the widespread and extensive variety of uses from the classification of images to activity acknowledgment, main challenge is detection, localization of the key points in the variation of body poses. To resolve this issue, substantial research work has been done in this area. After analyzing several results and detecting the restrictions, the author has reconstructed a simple model using neural network that estimates the poses and demonstrates the potential of CNN's. The author concludes with a few promising bearings and directions that have to be explored for future research.

[4] We propose to use a feature selection method that finds the relevant features for the learning task at hand using feature interaction (based on word interdependencies). This will allow us to reduce considerably the number of selected features from which to learn, making our KNN algorithm applicable in contexts where both the volume of documents and also the size of the vocabulary are high, like with the World Wide Web. [WWW].

[5] In this paper, different methods of text to speech synthesizer techniques are discussed to produce intelligible and natural output and a vowel synthesizer using cascade formant technique is implemented. A text to speech output is totally based on generating corresponding sound output when the text

is inputted. Wide range of applications use text to speech technique in many fields. The Various speech synthesis methods that have been used for text to speech output for obtaining intelligible and natural output are Concatenation, Informant, Articulatory, Hidden Markov model (HMM).

[6] Conventional approaches to statistical parametric speech synthesis typically use decision tree-clustered context-dependent hidden Markov models (HMMs) to represent probability densities of speech parameters given texts. Speech parameters are generated from the probability densities to maximize the output probabilities and then a speech waveform is re-constructed from the generated parameters.

3. Problem Definition

1. Wrongly stretching one's body or doing incorrect asanas, can be detrimental to your health. 2. Incorrect postures could lead to acute pain and long-standing chronic problems, say doctors. 3. Most yoga injuries develop gradually because of poor yoga forms or overdoing certain asanas. 4. Availability of Yoga Gurus in most part of the world is a big Challenge even in today's Era. 5. It is shocking to see the ignorance levels in Indian Yogis about the specific side effects of Yoga when not performed to perfection. 6. In fact each and every Asana can be really counterproductive if not done properly 7. There are many poses that are part of Simple Yoga - but not done with precision, can lead to hernia / sphincter damage and many many more.

4. Implemented Solution

1. Human Pose Estimation is defined as the problem of localization of human joints (also known as key points - elbows, wrists, etc.) in images or videos. 2. It is also defined as the search for a specific pose in space of all articulated poses. 3. YogAchar-E is smart platform which uses the above-mentioned technology to provide AI personal trainer to guide and correct Yoga position. 4. Pose Estimation based yoga trainer uses pose estimation to find reference key points on the body. These points are used to understand classify common Yoga positions. 5. Users interact with Pose Estimation based yoga trainer through both visual feedbacks via the mirror display and the voice interface using Snips AIR voice assistant. It enables the user to give commands start, stop, pause and restart Yoga. 6. YogAchar-E also talk-back through Flite voice synthesizer to guide the yogi to achieve the correct pose.

A. Block diagram

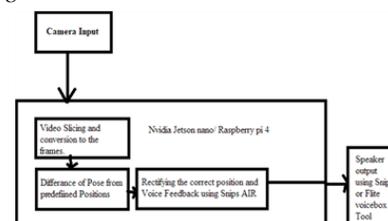


Fig. 1. Block Diagram of pose estimation

The above figure shows a pictorial block diagram of Pose Estimate Based Yoga Instructor. The figure consists of three major blocks namely camera input the speaker output find the

processing Centre which includes Nvidia Jetson Nano or Raspberry Pi 4. The camera input will provide, pictorial input of the Yog asanas to the Raspberry Pi or Nvidia Jetson which will slice the video feed and convert it into the input frames. These reference frames will then be passed through pose estimation algorithm which will compare these frames the predefined positions, the comparison will be made using KNN algorithm. Further in future advancement, it will even rectify the positions with a voice feedback using snip AIR and speaker that will use flight voice box tool.

5. Hardware Specification

A. Raspberry Pi4

The Raspberry Pi Pico uses RP2040 running at 133 MHz. Raspberry Pi will be the main Computer to Compute and Process the incoming feed with the reference Images and determine the Equivalent response.

B. NVIDIA Jetson Nano

NVIDIA Jetson will be the main Computer to Compute and Process the incoming feed with the reference Images and determine the Equivalent response.

C. Ada-fruit Microphone Amplifier Breakout

The Microphone Amplifier is an Electret Capsule Microphone with a MAX4466 Operational Amplifier optimized for use in microphone pre-amplifier applications. Microphone will be the Ear or the Audio sensor of our Project, any command given by the user will be processed into the electronic signals for further processing.

D. Raspberry Pi, Camera Module CSI

Raspberry Pi Camera will be the Optical Sensor of the project that will Feed the computer with the Visual/Image input that the user is providing, for further processing.

E. Flite (festival-lite)

It is a small, fast, open-source text to speech synthesis engine developed at CMU (Carnegie Mellon University) and primarily designed for small embedded machines and/or large servers.

F. Tensor-Flow

It is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular on deep neural networks. Here It will be the main program to do all the processing and Complex calculations.

G. Snips

Snips AIR will Add a voice to our Project by giving simple output response. In future Updates it will be responsible for giving Complex Feed-backs to correct the positions of the user

6. Future Work

- In this project our future aim is to extend the features so that Technology can be further used to teach Ramp walk and even Dance as well as in military applications.
- This technology can be used for Predictive analysis of Potential Armed Threats just by analyzing walking pattern

and stance.

- Also, for future our aim is to make the system as accurate as possible. As accuracy is the most important part of this project.

7. Advantages

1. Need for personal Yog Gurus will be minimized.
2. Precise and personal guidance will be provided by the AI.
3. In further updates user can ask questions to Pose Estimation based Yoga Trainer about health and fitness and it will be connected to internet hence assuring vast downloadable library of Yogas.
4. Personal attention and at the convenience solitude and oneness will be provided. Group sessions will be a thing of past. Hence contributing to avoid any gathering during this Pandemic.
5. AI would have a low error rate compared to humans, if coded properly. They would have incredible precision, accuracy, and speed.
6. Fewer errors.

Acknowledgment

I am profoundly grateful to Prof. Aarti B. for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. I would like to express deepest appreciation towards Dr. Vilas Nitaware, Principal Dr. Baban U. Rindhe HOD of EXTC Technology Department and Ms. Sushma Kore (Project Coordinator) whose invaluable guidance supported me in completing this project. At last I must express my sincere heartfelt gratitude to all the staff members of Information Technology Department who helped me directly or indirectly during this course of work.

8. Conclusion

This paper presented an overview on pose estimate based yoga instructor

References

- [1] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607-626, April 2009.
- [2] Y. Liu, Y. Xu and S. Li, "2-D Human Pose Estimation from Images Based on Deep Learning: A Review," 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, pp. 462-465, 2018.
- [3] A. Singh, S. Agarwal, P. Nagrah, A. Saxena and N. Thakur, "Human Pose Estimation Using Convolutional Neural Networks," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, pp. 946-952, 2019.
- [4] A. Acero, "An overview of text-to-speech synthesis," 2000 IEEE Workshop on Speech Coding. Proceedings. Meeting the Challenges of the New Millennium (Cat. No.00EX421), Delavan, WI, USA, 2000, pp. 1-4.
- [5] S. Lukose and S. S. Upadhyay, "Text to speech synthesizer-formant synthesis," 2017 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, 2017, pp. 1-4.
- [6] H. Ze, A. Senior and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, pp. 7962-7966, 2013.

- [7] <https://learn.adafruit.com/speech-synthesis-on-the-raspberry-pi/speak-easier-flite>
- [8] <https://paperswithcode.com/task/pose-estimation>
- [9] <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering>
- [10] <https://www.academia.edu/Documents/in/Pose/estimation>
- [11] <https://static.googleusercontent.com/media/research.google.com/en/pubs/archive/42237.pdf>