BENGAL COLLEGE OF
ENGINEERING & TECHNOLOGY
Approved by AICTE, New Delhi and Affiliated to MAKAUT, Kolkata
ISO 9001 : 2008 Certified Institute

Department of Computer Science and Engineering,
Bengal College of Engineering and Technology, Durgapur
(DA-8, Sahid Sukumar Banerjee Sarani, Bidhannagar, Durgapur, West Bengal 713212)

Human Body Posture Recognition

Final Year Project Report

Completed by -

Aniket Jha

Debangan Roy

Kunal Dhibar

Under the Guidance of Debashree Sarkar (Assistant Professor)

HUMAN BODY POSTURE RECOGNITION

This is submitted to the Department of Computer ***Science & Engineering/Information Technology*** of ***BENGAL COLLEGE OF ENGINEERING & TECHNOLOGY*** affiliated to ***Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of technology)*** for the requirements of the partial fulfillment of the Degree of ***Bachelor of Technology*** (***B. TECH) in Computer Science & Engineering/Information Technology.***

PROJECT SUBMITTED BY:

1. Aniket Jha (12500118110)
2. Debangan Roy (12500118094)
3. Kunal Dhibar (12500118079)

UNDER THE GUIDANCE OF:

Debashree Sarkar
Assistant Professor
Department of CSE/IT

BENGAL COLLEGE OF ENGINEERING & TECHNOLOGY

DURGAPUR - 713212

Certificate

This is to certify that this project *"**Human Posture Recognition**"* is a bonafide record of work done by:

1. **Aniket Jha (12500118110)**
2. **Debangan Roy (12500118094)**
3. **Kunal Dhibar (12500118079)**

This is submitted in partial fulfillment for the requirement of the degree of ***Bachelor of Technology*** in ***Computer Science & Engineering/Information Technology*** at Bengal College of Engineering & Technology affiliated to Maulana Abul Kalam Azad University of Technology, West Bengal.

Debashree Sarkar	**Sk. Abdul Rahim**	**Dr. P. K. Prasad**
Assistant Professor	Head of Department	Principal
Computer Science Engineering	Computer Science Engineering	BCET, Durgapur
BCET, Durgapur	BCET, Durgapur	

Acknowledgments

"It is not possible to prepare a project synopsis without the assistance & encouragement of other people. This one is certainly no exception." On the very outset of this paper, we would like to extend our sincere & heartfelt obligation towards all the personages who have helped us in this endeavor. Without their active guidance, help, cooperation & encouragement, we could not have made headway in the project.

First and foremost, we would like to express our sincere gratitude to our guide and mentor, Mrs. Debashree Sarkar. we were privileged to experience a sustained enthusiastic and involved interest from her side. This fueled our enthusiasm even further and encouraged us to boldly step into what was a totally dark and unexplored expanse before us. She always fueled our thoughts to think broad and out of the box.

We would also like to thank HOD (CSE) sir, Sk. Abdul Rahim who, instead of his busy schedule, always guided us in right direction. I would like to thank the entire staff member for motivation guidance and support. We are indebted to a number of friends and well-wishers who have extended their co-operation and help in the preparation of the project.

Last but not the least, it goes without saying that we are deeply indebted to our parents for their support and their patient guidance.

Thanking You

Debangan Roy   Kunal Dhibar   Aniket Jha

Table of Contents:

Abstract

The purpose of this paper is to describe about Machine Learning and Deep Learning using Python. In the modern-day python is becoming increasingly prevalent, thus it is important to know more about the language and its usage in Advance Computer Vision.

There are a variety of computer applications that identify human body in digital images, like - pedestrian crossing, criminal identification, healthcare and so on. The detection program allows us to identify and locate objects. It is very important in area of research, where the detected object can be count, accurately determined. The Python OpenCV library functions are mainly aimed at real-time computer vision.

The proposed system is able to efficiently model human postures by exploiting the depth information captured by an RGB-D camera. Firstly, a skeleton model is used to represent the current pose. Human skeleton configuration is then analyzed in the 3D space to compute joint-based features.

With the use of OpenCV, Media Pipe and other Machine Learning and Deep Learning Libraries Human Body and its various parts and joints are scanned for Body posture with their angles to estimate or evaluate the percentage of perfectness a user / person performs while doing a particular activity.

This paper will examine the benefits of Python in necessary utilizing technology and Computer Science in order to reach a variety of learners, and the versatile tactics Students, professors and Engineers from all around the world must employ to ensure its effectiveness.

Introduction

2.1. Overview

Throughout the global health is a concerned topic for everyone, thus everyone is immersed in keeping himself/herself healthy by doing different types of activities. With great use of the gym and different types of yoga, many different types of pains and issues arise among those performers. These generally occur due to maintaining the wrong posture while working out or doing their yoga. Also, it is found while sitting for a long time on chairs, sofas, etc. pains and a variety of issues are formed due to carrying out the activities in wrong positions or postures.

Here comes our project which helps the user to find the right posture. Generally, just by viewing or reading about the right Gym / Yoga postures, one can't carry out the same as shown in their source. Thus, our model (program) takes their saved or live feed video as input and provides a percentage as the output of how accurate their done posture for that particular workout is.

Our feature set characterizes the spatial configuration of the body through the 3D joint pairwise distances and the geometrical angles defined by the body segments using Media pipe library [1]. Posture recognition is then performed through a supervised classification method. In recent years, the use of skeleton data for human posture recognition has emerged as a popular research topic in the computer vision field. This technology shows good prospects for application in human-computer interaction, rehabilitation medicine, multimedia applications, virtual reality, robot control, and others. In general, postures are different from actions, with the former being static and the latter dynamic. A human posture is a base of actions, and is often taken as the key frame in various action recognition algorithms.

Moreover, in some fields, such as physical training, rehabilitation training and sign language communication, a human posture is more important than an action. In noisy workshops and dangerous working environments, posture recognition, as a human-computer interaction mode, is much superior to keystroke control and voice interaction in that it is more accurate, efficient and more natural in interaction.

2.2. Objectives

Human pose recovery and action recognition are one of the most fascinating topics in Computer vision. Several methods have been proposed for human pose recovery which uses a pictorial structural framework that revolutionized this area of research. For human pose recovery, we used local part-based models, and constrained are imposed on the joints geometrically. This model helps us deal with the problems of foreshortening and rotation of limbs.

Action recognition is one of the important topics in computer vision. It is important for the applications such as video surveillance, content-based video search. There are lots of other fields as well. Knowing the orientation of a person opens avenues for several real-life applications, some of which are discussed towards the end of this blog. Several approaches to Human Pose Estimation were introduced over the years. The earliest (and slowest) methods typically estimate the pose of a single person in an image that only had one person, to begin with. These methods often identify the individual parts first, followed by forming connections between them to create the pose. There are several main methods for posture recognition [2]. One is to use wearable sensors, such as wearing accelerometer and pressure sensor. However, wearing such a device makes subjects feel a sense of burden, which compromises the interactive experience. The other one is based on monocular cameras. However, it is susceptible to illumination and background interference, offering unsatisfactory recognition accuracy and robustness in complex conditions. With the increasingly low-cost depth image sensors, RGB-D image-based posture and action recognition has become an important research focus in the field of human-computer interaction. Researchers can obtain color and depth images as well as skeleton data of human easily [3]. Many posture recognition algorithms that use skeleton data obtained from Kinect are proposed. These algorithms can not only avoid the influence of illumination, but they also eliminate the need of preprocessing such as segmentation and object detection in complex backgrounds, which enables greatly improved accuracy. Additionally, datasets and algorithms based on posture recognition are still of limited availability. Therefore, in this paper we propose a human posture recognition method, (figure 1) which incorporates the use of pre-built dataset model i.e., Mediapipe by Google.

This will change the use of Machine Learning (Deep Learning) in future scope, implementing the Image Processing concept and giving us some wonderful results.

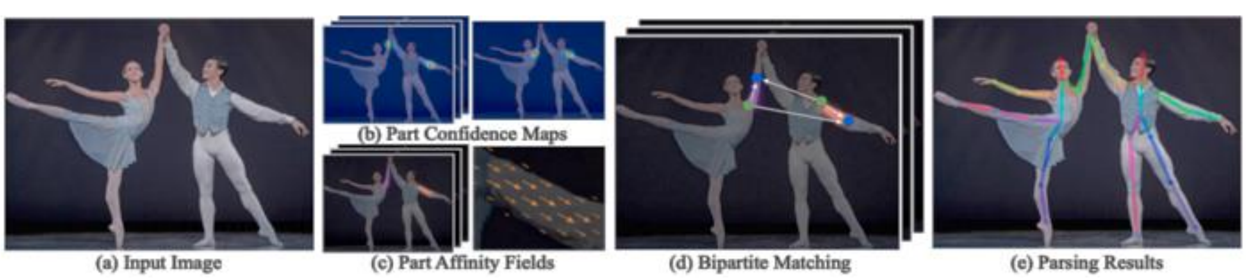


Figure (1): different layers of feature extraction from images.

Algorithms Used

3.1. Technologies used

Machine Learning: Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. Machine Learning is used in our project to make the model understand how the user is accurate to the given data (test data).

Deep Learning: Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. Deep Learning is used in our project to differentiate between various body parts like leg, forearm, elbow etc.

Neural Network: A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural Network is used to extract the valuable information like angle at which all joints move and estimate the posture.

Open CV: OpenCV Python is a library of Python bindings designed to solve computer vision problems. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection. OpenCV Python is nothing but a wrapper class for the original C++ library to be used with Python. Using this, all the OpenCV array structures gets converted to/from NumPy arrays. This makes it easier to integrate it with other libraries which use NumPy.

Media Pipe: Media Pipe is a cross-platform framework for building multimodal applied machine learning pipelines. Media Pipe is a framework for building multimodal (e.g., video, audio, any time series data), cross platform (i.e., Android, iOS, web, edge devices) applied ML pipelines.

Python: Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Report on Investigation

4.1. Basics

The code basically gives the 26 rectangular boxes which means there is one body part in each of the box. The model represents the human body as a deformable configuration (like spring is attached between parts) of individual parts which are in turn are modelled separately in a recursive manner. Then for all the configurations of human body parts, score is calculated as follows (from equation 1)

$$score^i(t_i,p_i)=b_i^{t_i}+w^i_{t_i}\bullet\varphi(I,p_i)+\sum_{k\in k_i d_s(i)} m_k(t_i,p_i)$$

equation (1)

Informally we can think of this equation as having two parts, one its local score (the 1st part of RHS) which means how much this part fits in the given position. And a pair-wise score which finds its score when their kids are at a fixed position. Now this score is propagated till its root (head) and the configuration which achieves the highest score is being selected. So, from this we get 26 parts of human body [4]. But our main goal is to recognize actions based on above data, so all the 26 parts individually doesn’t make much sense and increases the computation overhead as well. So, to minimize the cost we clustered the 26 body parts into 11 parts which are:

i. Left Hand/Arm/Torso/Thigh/Leg,

ii. Right Hand/Arm/Torso/Thigh/Leg, and,

iii. Head.

For clustering this part, we first normalized the skeleton w.r.t Head-Neck length. This method helps us deal with the cases of difference in height, length of detected skeleton caused by the different shapes and sizes of individuals. After normalization we used ‘Linear Regression’ method to estimate the 11 body parts described above.

After this we get a model which shows labels for each part of human body. Results obtained on some of the input images is as shown in the following figure (figure 2):

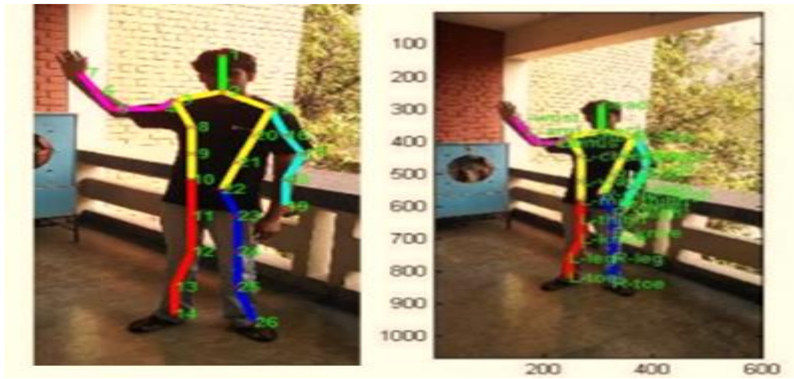


Figure (2):

4.2. Action Recognition

Action recognition from video Feature Extraction: From above 11 parts, we use the set of 8 angles as the feature vector of one image/frame. These angles are: Angles between: Left Hand and Arm, Left Arm and Torso, Left Torso and Left Thigh, Left Thigh and Left leg and similarly for Right body parts [5].

This provides a dataset of HD image sequences of 8 person (different camera view) doing 13 actions namely Walk, Run, Jump, Bend, Hand wave, Jump-in-place, Sit-Stand Up, Run-fall, Walk-sit, Run-jump walk, Handshake, Pull, Facial-expressions.

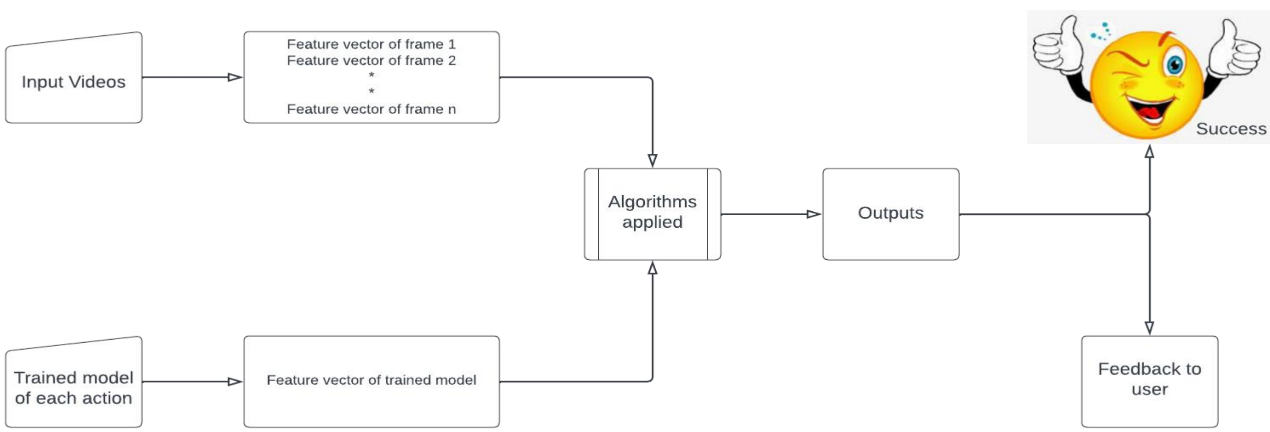


Figure (3): Flow chart to understand step by step working of model.

At the end of test video, we will ask the user whether the result is consistent with the provided video or not, if not we ask user the correct action of that video and train our model on this new video accordingly.

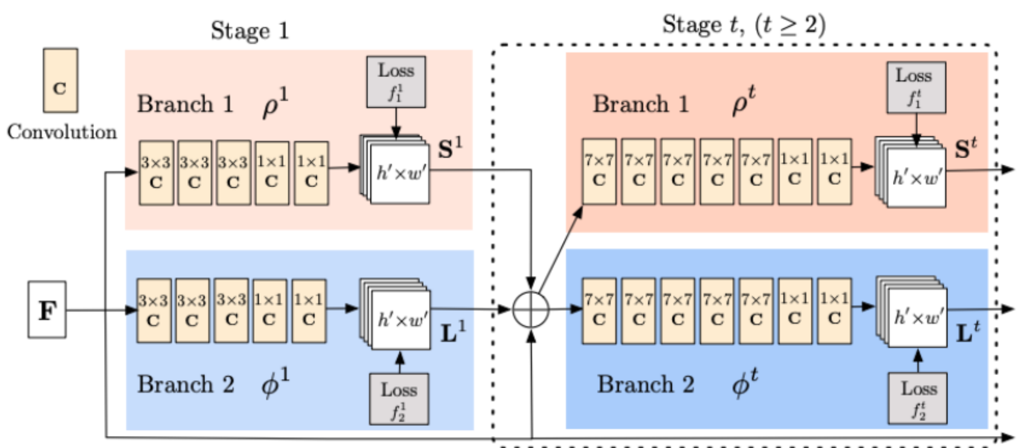
Methodology

5.1. Methods Proposed

a. **OpenPose -**

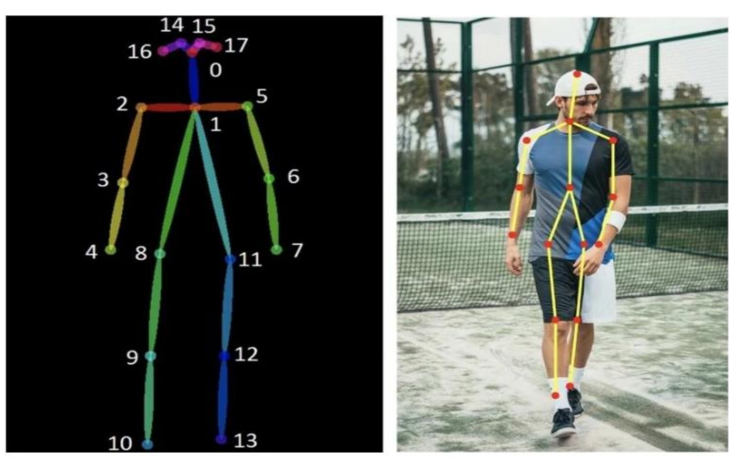
OpenPose is one of the most popular bottom-up approaches for multi-person human pose estimation, partly because of their well-documented GitHub implementation.

As with many bottom-up approaches, OpenPose first detects parts (key points) belonging to every person in the image, followed by assigning parts to distinct individuals. Shown below is the architecture of the OpenPose model [6].



OpenPose Data Flow Diagram

Figure (4)



OpenPose Diagram

Figure (5)

b. **DeepCut**

DeepCut is a bottom-up approach for multi-person human pose estimation. The authors approached the task by defining the following problems:

- Produce a set of body part candidates. This set represents all possible locations of body parts for every person in the image. Select a subset of body parts from the above set of body part candidates [7].
- Label each selected body part with one of the body part classes. The body part classes represent the types of parts, such as "arm", "leg", "torso" etc.
- Partition body parts that belong to the same person.

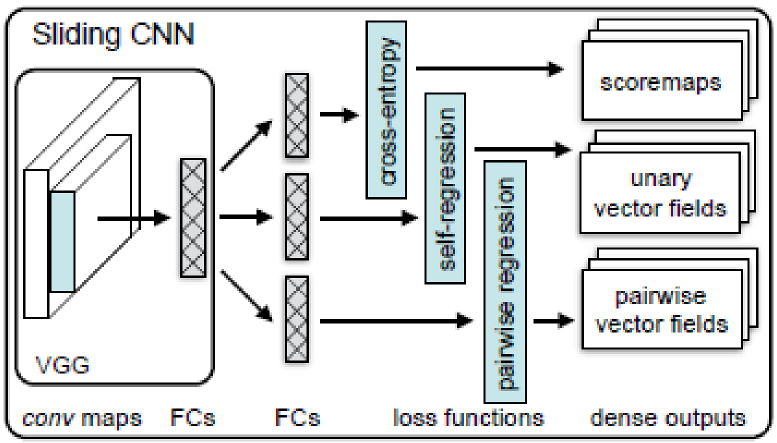


Figure (6)

DeepCut Architecture

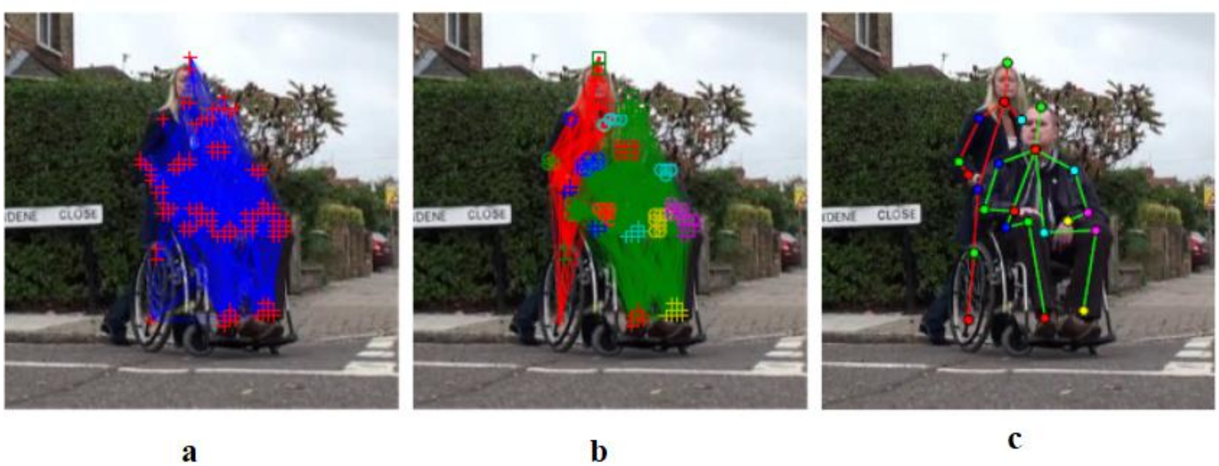


Figure (7)

a. Initial detections and pairwise terms (graph) between all detections.

b. Detections that are jointly clustered belonging to one person.

c. The predicted pose sticks

c. **RMPE (AlphaPose) -**

RMPE is a popular top-down method of Pose Estimation. The authors posit that top-down methods are usually dependent on the accuracy of the person detector, as pose estimation is performed on the region where the person is located. Hence, errors in localization and duplicate bounding box predictions can cause the pose extraction algorithm to perform sub-optimally [8].

Effect of duplicate predictions (left) and low confidence bounding boxes (right):

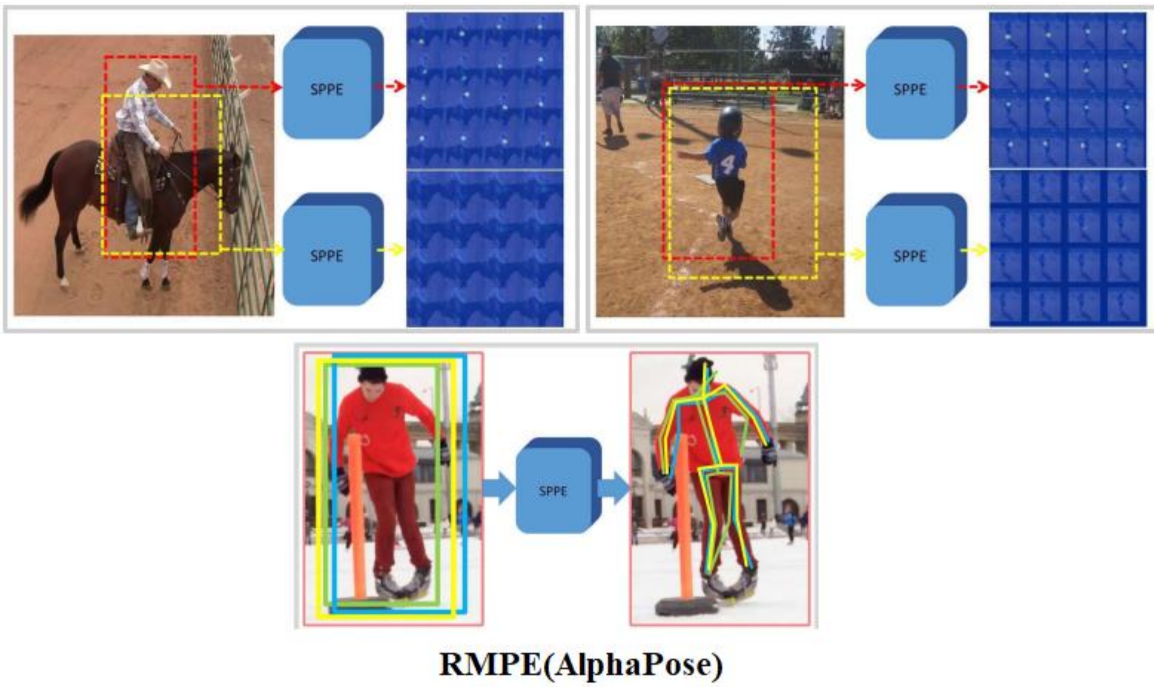


Figure (8)

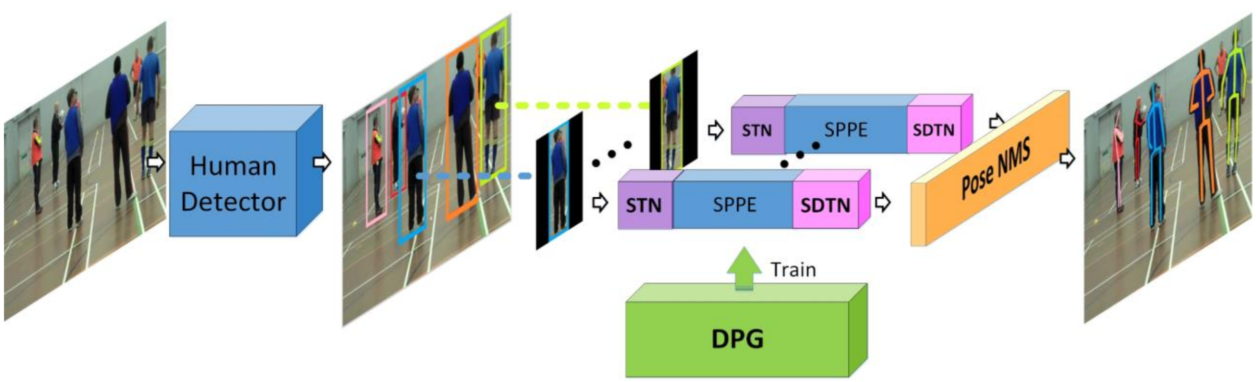


Figure (9)

d. **Mask RCNN -**

Mask RCNN is a popular architecture for performing semantic and instance segmentation. The model parallelly predicts both the bounding box locations of the various objects in the image and a mask that semantically segments the object. The basic architecture can be quite easily extended for human pose estimation [9].

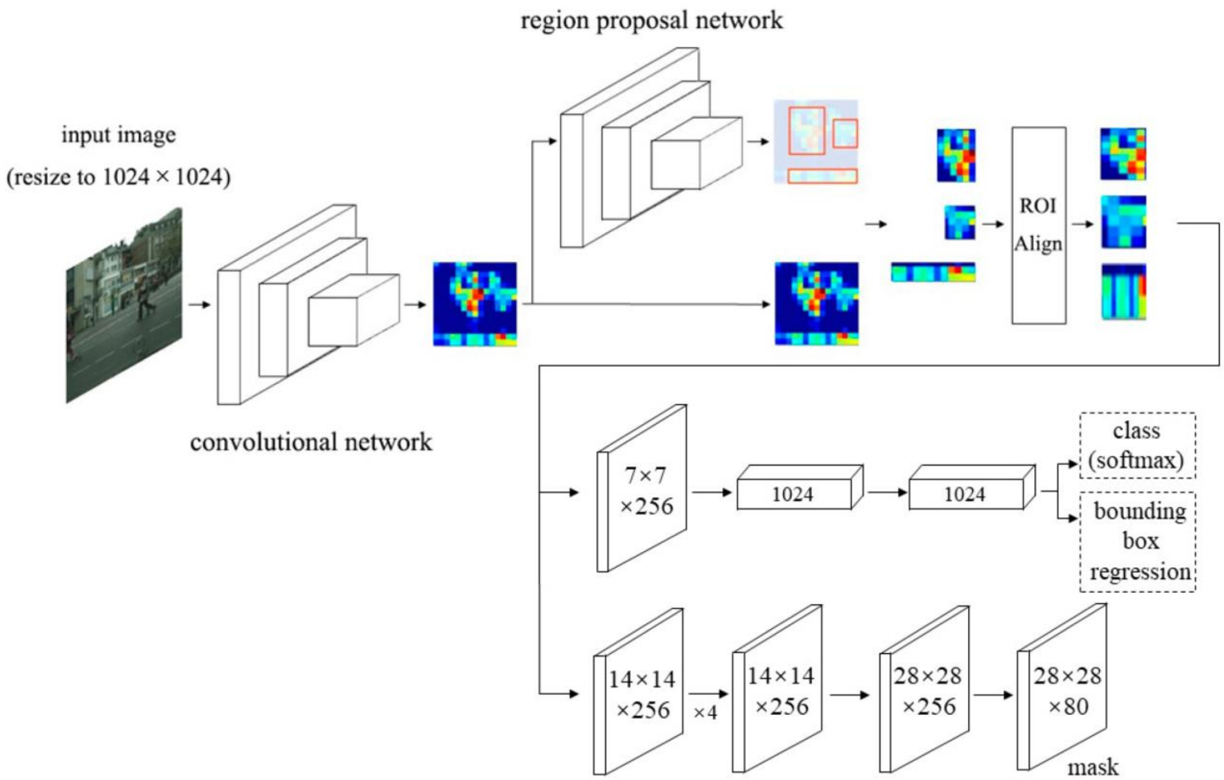


Figure (10): Flowchart of Mask RNN Architecture

5.2. Method Used

In this paper Open Pose approach is used. It is one of the best bottom-up approach to trace a human body and create the mesh.

The reason for choosing is because OpenPose is much more accurate in the case of 2-dimensional posture detection than other methods. Unlike DeepCut, it allows the user to take the specific parts of the body to do the desired job. OpenPose can also be used in both live processes and recorded processes. Also, it allows detecting multi-person systems [10].

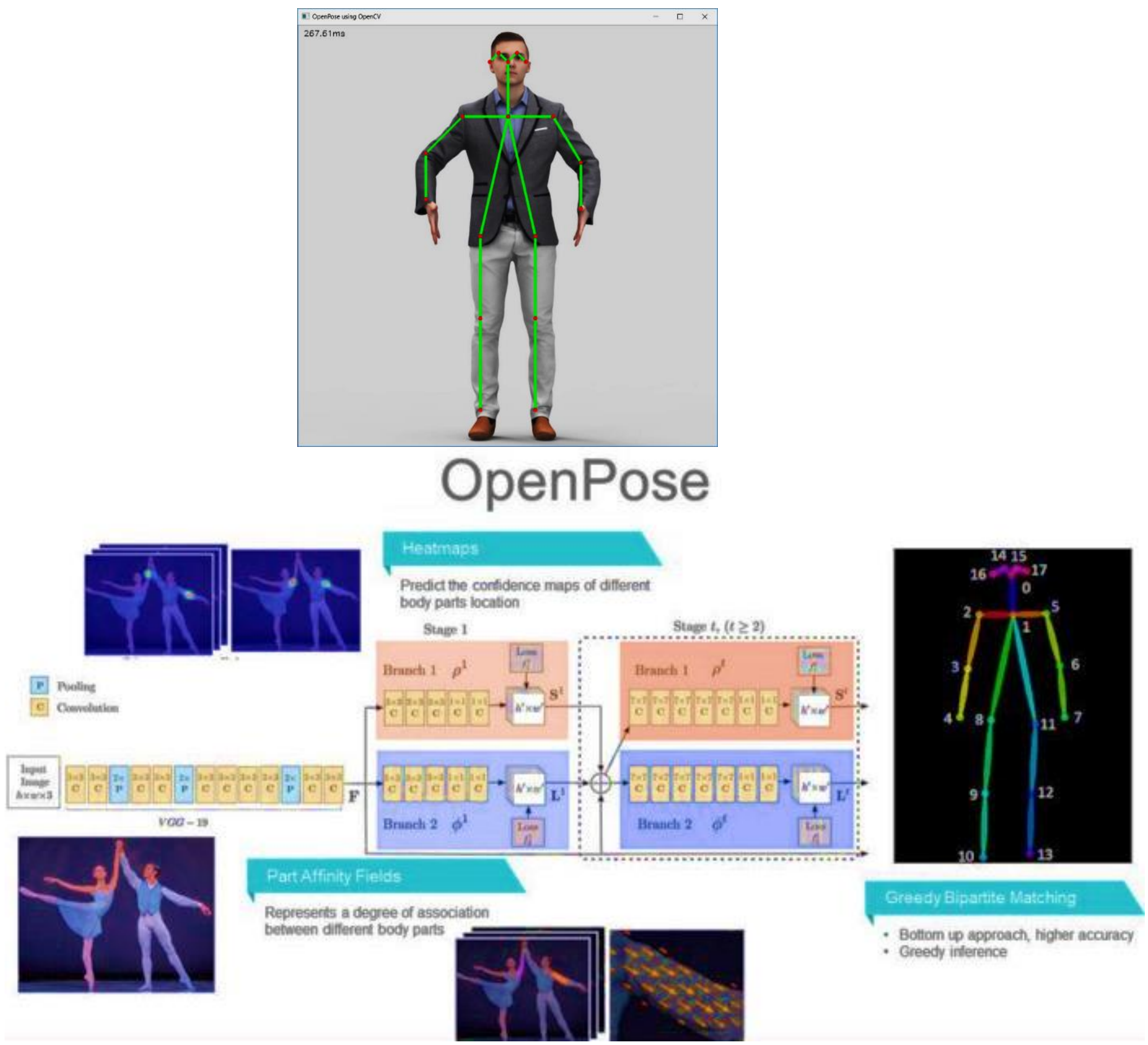
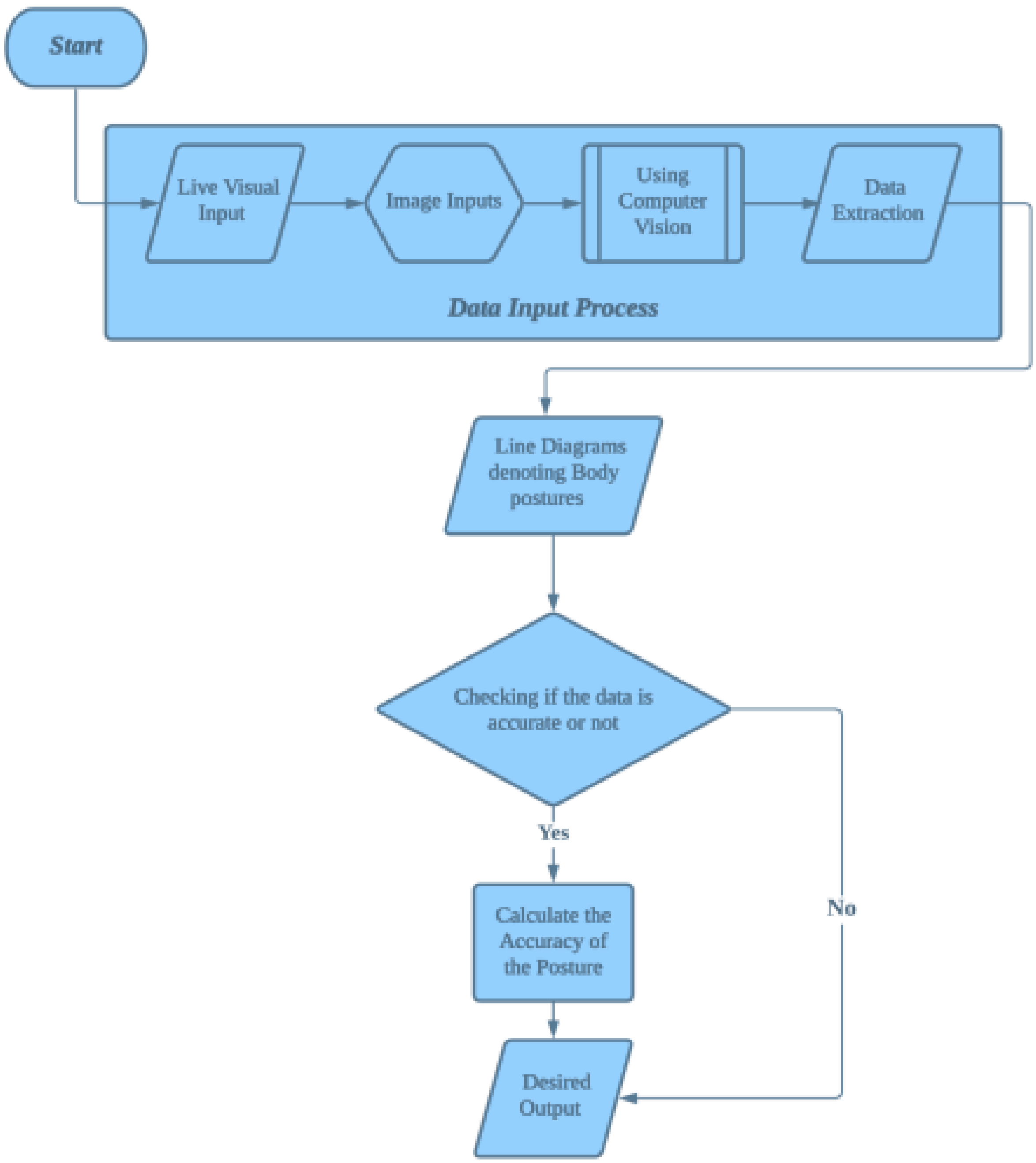


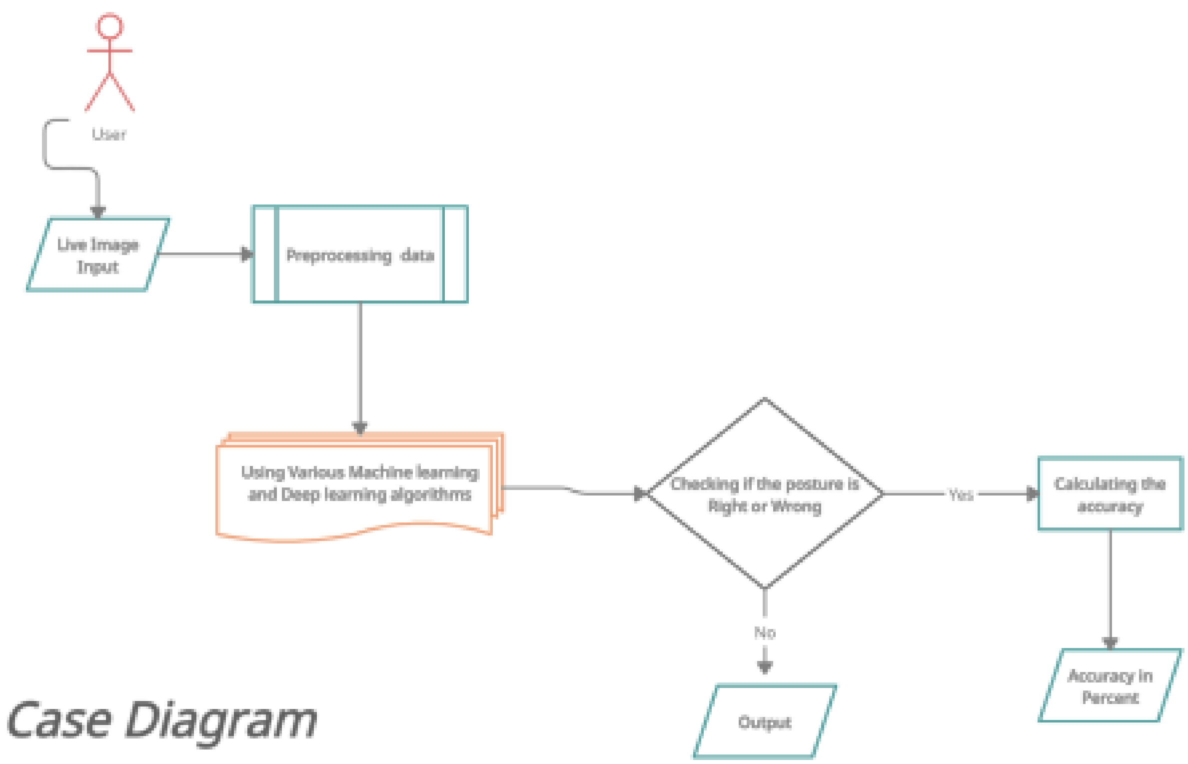
Figure (11) : Working of Open Pose.

Flowcharts and Diagrams

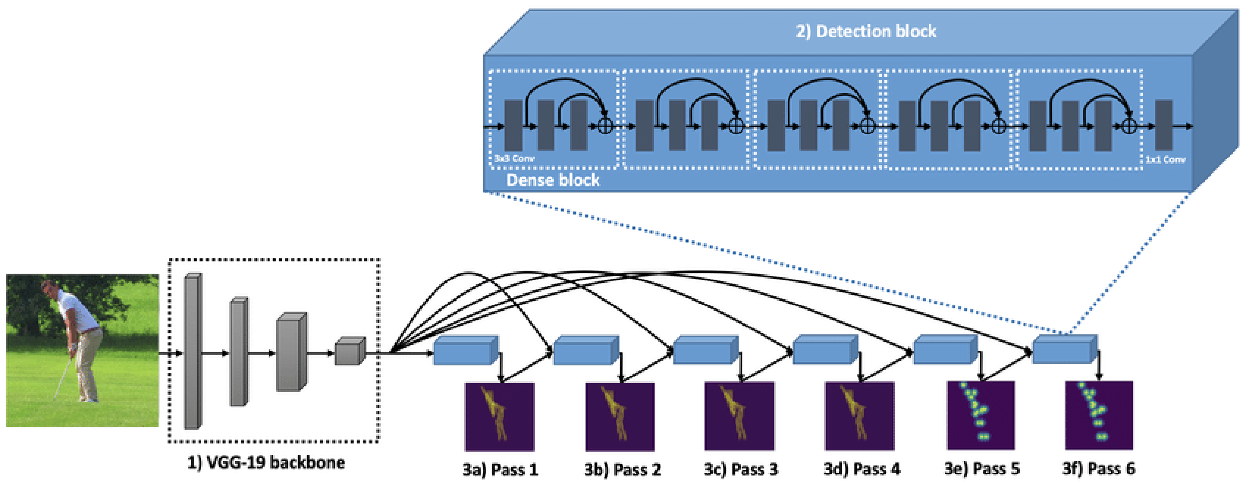
6.1. Data Flow Diagram

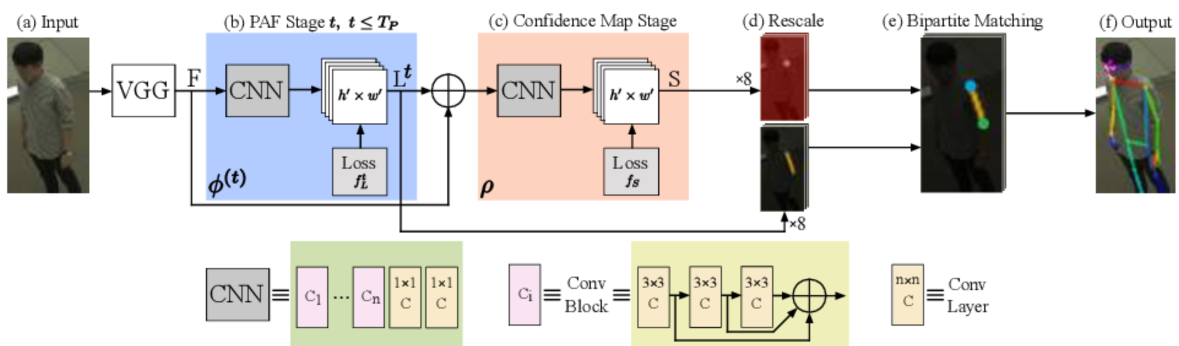


6.2. Case Diagram

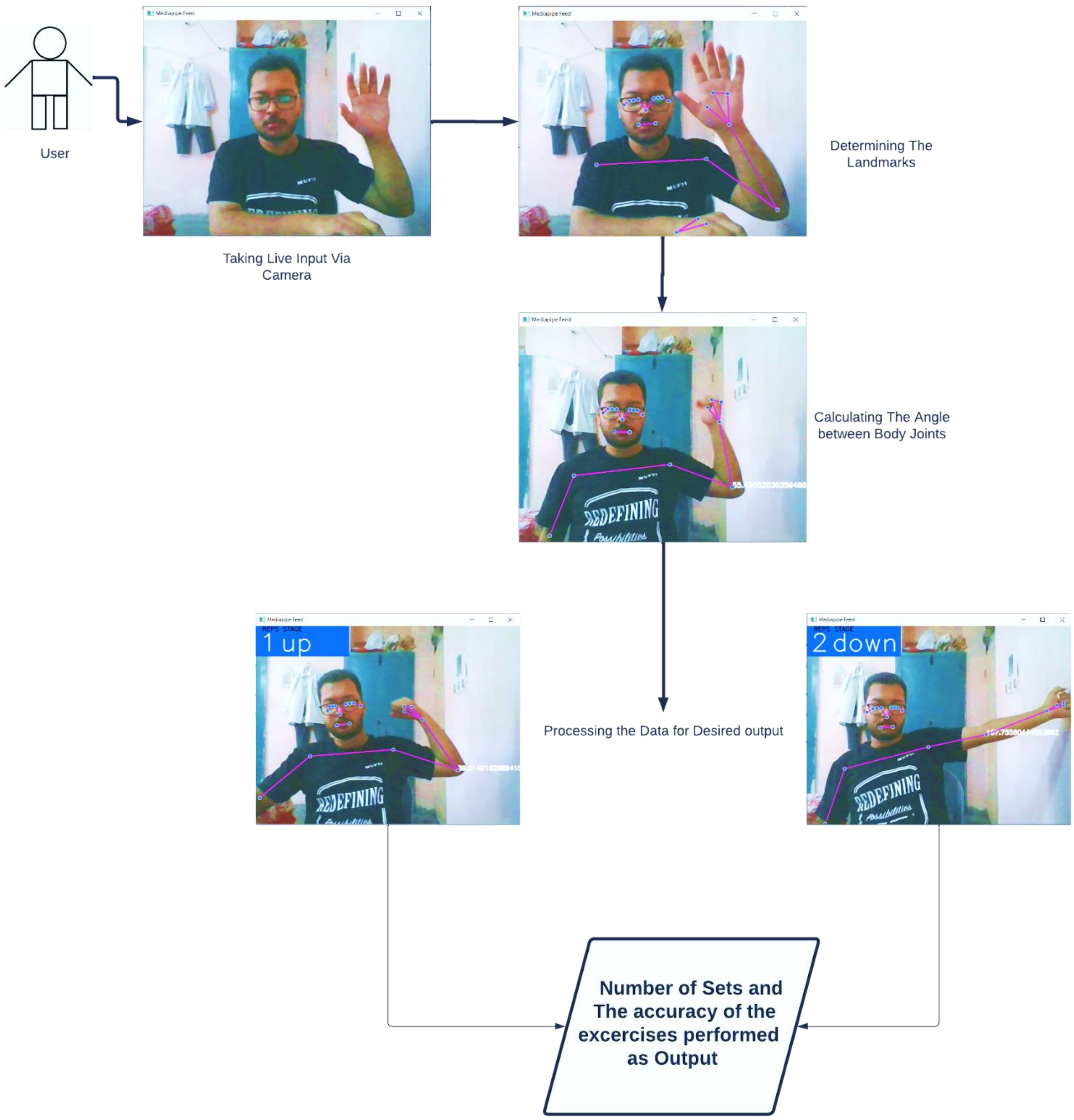


6.3. DataFlow of the Algorithm





6.4. Pictorial reference of DFDs



Result and Discussion

We tried to implement a good model for human action recognition-based skeleton information of the human body represented by an 8 - dimensional feature vector. We implemented our own algorithm to classify action into 5 categories and were also successful in it. More information about the process can be seen in the paper [11].

For more robust approach we also have to take the depth data in consideration and this can be implemented as a real time action classifier because of lesser computation cost in second part of our algorithm.

Implementation and Future Scope

The model can be accessed and used by running the source code in any IDE. This model can be used in Web, Android, or IOS applications.

This application can change the vision of animation, CGI and any other technologies which involves human body posture and its movement. It can be further used in robotics to train the machines.

Conclusion

From the above project we can detect the posture and angles of the body. Further this project will be used to determine the accuracy of the exercises (which we are working on), and finally converting it into a full-fledged application as an *"AI Personal Trainer"*.

References

[1] Oshima, Takashi. (2021). Image processing device, image processing method and image processing system.

[2] Bräunl, Thomas. (2022). Image Processing. 10.1007/978-981-16-0804-9_16.

[3] Nithya, M. & .V, Lalitha & Paveethra, K & Kumari, S. (2022). Early Detection of Parkinson's Disease using Machine Learning & Image Processing. 1-4. 10.1109/ICCCI54379.2022.9740961.

[4] Xie, Fei & Xu, Guili & Cheng, Yuehua & Tian, Yupeng. (2009). An improved thinning algorithm for human body recognition. 2009 IEEE International Workshop on Imaging Systems and Techniques, IST 2009 - Proceedings. 10.1109/IST.2009.5071678.

[5] Host, Kristina & Ivašić-Kos, Marina & Pobar, Miran. (2022). Action Recognition in Handball Scenes. 10.1007/978-3-030-80119-9_41.

[6] Han, Yifeng & Han, Rui & Liu, Li & Jingcheng, Mei & Qi, Donglian. (2019). A real-time video detection method of the surge arrester handling mode based on Yolo-v3 and Open-Pose. 8556-8561. 10.23919/ChiCC.2019.8865362.

[7] Jin, Qibing & Li, Ziming & Yan, Zhicheng & Wang, Bin & Wang, Zeyu. (2022). Optimization Study on Enhancing Deep-Cut Effect of the Vacuum Distillation Unit (VDU). Processes. 10. 359. 10.3390/pr10020359.

[8] Fang, Haoshu & Xie, Shuqin & Lu, Cewu. (2016). RMPE: Regional Multi-person Pose Estimation.

[9] S, Mohanapriya & P, Natesan & S, Mohana & P, Sumithra & M, Swathi & P, Yoga. (2022). Instance Segmentation using Mask RCNN for Surveillance. 235-241. 10.1109/ICCMC53470.2022.9753924.

[10] Shihab, Amanj & Noori, Zainab & Atiyah, Jasim. (2022). Open Source Platforms and Recognition for Image Processing. 12. 11-34.

[11] Rahman, Shahina & Johnson, Valen. (2018). A Fast Algorithm for Clustering High Dimensional Feature Vectors.