Human Pose Estimation Using Machine Learning

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

Anjali Nirav Shah, shahanjali0604@gmail.com

Under the Guidance of

P.Raja, Master Trainer, Edunet Foundation

ACKNOWLEDGEMENT

I am very grateful to all the individuals who have supported and guided us, both directly and indirectly, throughout the completion of the project as well as obliged

First and foremost, I would like to thank my supervisor and guide, P. Raja, for being a great mentor and the best adviser I could ever have. His unwavering support, invaluable guidance, and insightful critiques have been a source of innovative ideas and inspiration throughout this journey. Being a beginner to this project, he guided each and everything which was essential for the development of the project as well as confidence to complete the work. It has been a privilege working with him for the last one year. His mentorship has not only enhanced our project work but also contributed significantly to our professional growth. His teachings have shaped us into more responsible and capable professionals, ready to face future challenges.

I am equally obliged to have this amazing opportunity with the Microsoft, SAP-AICTE, TechSaksham and Edunet Founation which allowed us to expand our knowledge and develop skills critical to the field of AI.

Lastly, I would also like to express my gratitude to my family and friends for their unwavering support and understanding throughout this process. Their encouragement and patience have been my pillars of strength.

ABSTRACT

This project focuses on capturing human poses and analyzing their complexities through the implementation of fundamental human pose estimation techniques. It addresses the challenges associated with accurately detecting human postures despite occlusions, varying lighting conditions, and diverse body types.

The objectives include designing an efficient algorithm to identify key body points, minimizing detection errors, and improving overall system accuracy. The methodology involves using techniques such as OpenCV, Streamlit, and MediaPipe for image processing and analysis. A dataset of annotated human poses is used for training and validation. The model detects key body points, such as joints, and constructs a skeletal model representing the human pose.

Key results demonstrate that the proposed beginner model achieves satisfactory accuracy in detecting and estimating human poses under challenging conditions. The algorithm successfully addresses occlusions and varying lighting conditions, providing reliable pose estimation across different body types and postures. Evaluation metrics indicate promising improvements, with a reduction in error rates.

In conclusion, this project presents a foundational approach to human pose estimation, addressing common challenges in the field. A significant application of this model is in virtual body measurement for online clothing fittings, which can revolutionize the fashion industry by providing accurate size recommendations. Other practical applications include detecting abnormal movements for medical purposes and tracking players' movements in sports. Future work may involve expanding the dataset and further optimizing the algorithm for real-time applications.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

Human Pose Estimation, as the name implies, detects the position of the human body. This technology is crucial in scenarios where activities require precise measurement and analysis of body movements. The complexity arises in accurately detecting and interpreting these movements, which can include tracking ill practices, judging false movements, and ensuring precision in various applications. One of the primary challenges is the accurate detection of human postures despite factors like occlusions, diverse body types, and varying lighting conditions.

This project aims to address these challenges by developing a model capable of estimating body positions with high precision across different scenarios. The model simplifies the process of detecting, tracing, and judging specific body movements in various fields of societal work. Whether it's in sports, where tracking player positions and identifying critical movements can help judges make fair decisions, or in healthcare, where correcting postures can prevent injuries during activities like dance and yoga, the applications are vast.

Additionally, this pose estimation technology can be instrumental in business environments for identifying unethical practices. It can also benefit designers by providing accurate body measurements for customized clothing, enhancing the user experience in online shopping. By focusing on these diverse applications, the project seeks to provide a robust solution to the complex problem of human pose estimation, offering significant benefits across multiple domains.

1.2 Motivation:

This project detects the position of the human body and structures the use of artificial intelligence (AI) to train the model according to the use, achieving greater precision. The main motivation behind building this model for estimating human positions lies in its various scopes and applications. Accurate detection can measure body lengths and various dimensions, which is beneficial for many people. This capability makes the project relevant and impactful in numerous domains, providing valuable assistance to a wide range of individuals.

The various applications and impacts include tracking the positions of players in sports, especially in critical situations, to help judges capture precise movements and encourage fair play. This technology aids users in correcting their postures to prevent injuries during activities such as dance, yoga, and other similar pursuits. By reducing the risk of injuries, the model helps mitigate medical emergencies and accidents to some extent. In business activities, it can trace unethical practices in specific activities or environments, enhancing security and compliance.

Furthermore, designers can utilize this model to offer customized features for their clients, enabling accurate body measurements for tailoring outfits. This application can revolutionize the fashion industry by providing precise and personalized clothing fittings. Overall, the motivation behind this project is to harness AI's potential to improve accuracy, efficiency, and safety across various fields, ultimately benefiting society at large.

1.3Objective:

The primary objective of this project is to develop a precise and reliable model for human pose estimation, leveraging artificial intelligence to achieve greater accuracy. This involves designing an efficient algorithm to identify key body points, such as joints, and constructing a detailed skeletal model of the human pose. The goals include reducing detection errors and improving the overall accuracy of the system by utilizing a trained model.

Additionally, the project aims to explore practical applications of pose estimation, such as tracking players' positions in sports to help judges make fair decisions and aiding users in correcting their postures to prevent injuries during activities like dance and yoga. The model will also be utilized for detecting unethical practices in business environments and offering customized features for designers to provide accurate body measurements for their clients.

By employing tools like OpenCV, Streamlit, and MediaPipe, the project seeks to ensure user-friendly implementation and usability in various scenarios. Ultimately, the objective is to harness AI's potential, through the trained model, to improve accuracy, efficiency, and safety across multiple domains, providing significant benefits to society.

1.4 Scope of the Project:

Scope

The project aims to develop a precise and reliable model for human pose estimation using artificial intelligence. This involves designing an algorithm to accurately identify key body points and constructing detailed skeletal models. The model will be trained with a diverse dataset to ensure high accuracy across different body types and conditions. One significant application is the virtual body measurement system, which offers precise size recommendations for online clothing fittings. The goal is to provide valuable assistance in sports, healthcare, and business environments.

Additionally, the model aims to simplify the detection and tracking of body movements in various activities, such as sports, where it can help judges capture critical movements and ensure fair play. For healthcare, it assists users in correcting postures to prevent injuries during activities like dance and yoga. In business, it can detect unethical practices, and for designers, it offers customized features to provide accurate body measurements for tailoring outfits. Overall, the project seeks to improve accuracy, efficiency, and safety across multiple domains, benefiting a wide range of users.

Limitation

While the project strives for high precision, it faces certain limitations. The accuracy of the model is influenced by the diversity and quality of the training dataset. Environmental factors like lighting and occlusions can affect performance. High computational resources are required for training and real-time estimation, which may not be feasible for all users. The model may struggle to generalize to unseen poses or activities not part of the training data. Achieving real-time performance with high accuracy can be challenging, especially on devices with limited processing power. These factors need consideration for the model's optimal application.

CHAPTER 2

Literature Survey

2.1 Relevant Literature or Previous Work in This Domain

Human pose estimation has been a significant area of research within computer vision, with numerous studies focusing on improving the accuracy and robustness of pose estimation models.[1]
Early methods relied on traditional manual modeling techniques, but the advent of deep learning has revolutionized the field.[2]

Recent literature, such as the systematic review by Samkari et al. (2023), highlights the use of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for enhancing pose estimation performance.[3]

Challenges such as occlusions, varying lighting conditions, and complex backgrounds remain prevalent issues.

2.2 Existing models, techniques, or methodologies related to the problem.

Several models and techniques have been developed to tackle human pose estimation.
OpenPose is a real-time system that detects human body, hand, and facial keypoints, while HRNet maintains high-resolution representations throughout the process, enhancing accuracy.
PoseNet offers a lightweight solution for real-time applications on mobile devices, and AlphaPose combines bottom-up and top-down approaches for robust multi-stage pose estimation.
DeepCut utilizes deep learning for effective multi-person pose estimation.
Despite these advancements, achieving centimeter-level accuracy remains a challenge, particularly in complex scenarios.

2.3 The gaps or limitations in existing solutions and how your project will address them.

Despite the advancements in human pose estimation, several gaps remain in achieving centimeter-level accuracy for virtual body measurement. Current models struggle with precision, particularly in complex scenarios involving occlusions and crowded scenes. Real-time processing on edge devices is another significant challenge due to the computational demands of sophisticated algorithms. Moreover, many models do not generalize well across different datasets and environments, which limits their applicability in diverse settings.

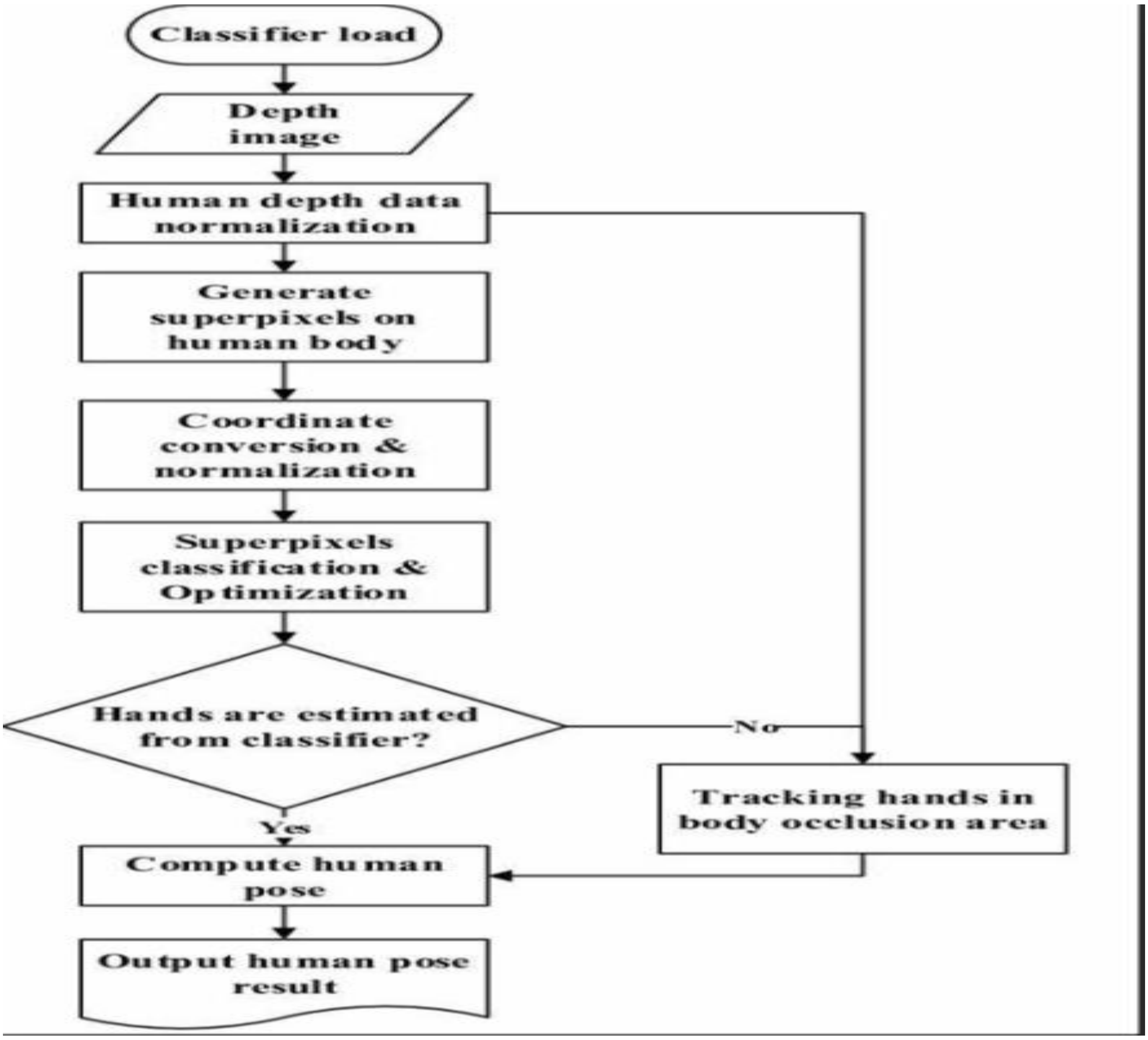
These gaps can be addressed by developing advanced algorithms capable of handling variations in body shapes, poses, and clothing more effectively. By integrating high-resolution imaging and multi-view systems with depth sensors, it is possible to provide more detailed and accurate body measurements. Additionally, optimizing for real-time processing and addressing ethical concerns related to user privacy and data security will ensure the responsible and fair use of the technology. This comprehensive approach will significantly enhance the accuracy and robustness of virtual body measurement using human pose estimation.

CHAPTER 3

Proposed Methodology

3.1 System Design

Figure 1



The flowchart represents the step-by-step process used in a human pose estimation system. Each block in the flowchart corresponds to a specific stage in the system's pipeline, from capturing a depth image to outputting the final human pose.

1. Classifier Load:
   a) The first step in the human pose estimation system is to load a pre-trained classifier model.

b) This classifier is designed to detect various parts of the human body within depth images, providing the foundation for the subsequent analysis.
c) By loading this model, the system ensures it has the necessary tool to identify and classify different body regions accurately.
d) This initial step is crucial, as it sets the stage for the entire pose estimation process, ensuring the system is equipped to recognize body parts effectively.

2. Depth Image:
   a) The next step involves capturing a depth image of the human subject.
   b) This image provides crucial depth information, indicating the distance of each point from the camera.
   c) By gathering this raw data, the system lays the groundwork for further processing and analysis.
   d) The depth image is essential for accurately modeling the spatial relationships of the human body in three dimensions.

3. Human Depth Data Normalization

After capturing the depth image, the system normalizes the depth data to ensure it is consistent and suitable for processing. This step adjusts variations in depth data, making it easier to analyze accurately. Normalization is vital for maintaining the integrity of the data and ensuring reliable results in subsequent steps.

4. Generate Superpixels on Human Body

The system then segments the depth image into smaller, homogeneous regions known as super pixels. These super pixels help in simplifying the image and enhancing processing efficiency. By breaking down the image into manageable segments, the system can more easily identify and classify different parts of the human body.

5. Coordinate Conversion & Normalization

Next, the coordinates of the super pixels are converted and normalized to maintain consistency. This step ensures that the spatial data is accurately represented and ready for classification. Proper coordinate conversion and normalization are crucial for preserving the relationships between different regions of the image.

Super pixels Classification & Optimization

The pre-trained classifier is used to classify the super pixels, identifying various body parts. Optimization techniques are applied to improve accuracy and refine the classification process. This step is critical for accurately identifying and categorizing different regions of the body.

Hands Are Estimated from Classifier?

At this stage, the system checks whether the hands are correctly estimated by the classifier. This step is a decision point that determines the next course of action based on the success of hand estimation.

If Yes:

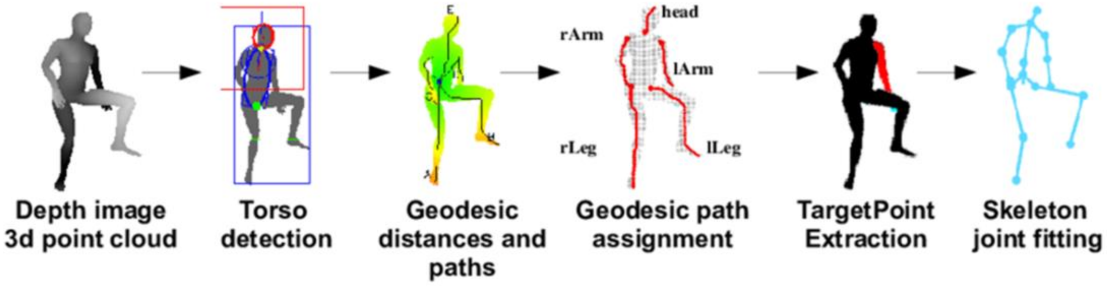
Compute Human Pose: The system uses the classified superpixels and their coordinates to compute the human pose.

Output Human Pose Result: The final estimated human pose is output.

If No:

Track Hands in Body Occlusion Area: The system tracks the hands in areas where the body may occlude them.

Loop Back to Superpixels Classification & Optimization: The process loops back to reclassify and optimize until accurate hand estimation is achieved.



Working Of Human Pose Estimation

Working:

Depth Image 3D Point Cloud

In this initial step, a depth image of the human subject is captured. This image provides depth information for each pixel, indicating how far each point is from the camera. The objective is to convert this depth image into a 3D point cloud, which consists of spatial coordinates (x, y, z) representing points on the surface of the human subject. This 3D point cloud serves as the foundation for further processing in human pose estimation.

Torso Detection

Once the 3D point cloud is generated, the next step is to identify the torso region within it. The torso is a key part of the human body, and detecting it accurately is crucial for further analysis. The system highlights the torso with bounding boxes and circles to isolate this region, making it easier to focus on and analyze the torso separately from other body parts.

Geodesic Distances and Paths

In this step, the system calculates geodesic distances on the surface of the 3D point cloud. Geodesic paths represent the shortest path along the surface between two points. These distances and paths are visualized using a color gradient, where different colors indicate varying distances from a reference point. This visualization helps in understanding the spatial relationships on the surface of the human body.

Geodesic Path Assignment

After calculating the geodesic distances, the system assigns geodesic paths to different body parts such as the head, arms, and legs. By marking these paths on the 3D point cloud, the system facilitates accurate identification and segmentation of various body parts. This step is essential for distinguishing and analyzing different regions of the human body.

Target Point Extraction

In this step, key points, also known as target points, are extracted from the geodesic paths. These target points are critical for identifying key joints and segments of the human body. Extracting these points enables the system to pinpoint important locations on the body, which are necessary for building a skeleton model.

Skeleton Joint Fitting

The final step involves fitting a skeleton model to the extracted target points. The skeleton model consists of connected joints, forming a simplified representation of the human pose. By aligning the skeleton with the target points, the system creates a coherent and accurate representation of the human body's structure and pose, which can be used for various applications such as motion capture and augmented reality.

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

The hardware requirements are as listed as below:

i. RAM: At least 16GB of RAM is recommended, but 32GB or more is ideal for handling large datasets and ensuring smooth performance.
ii. Camera: High-resolution cameras (1080p or higher)
iii. Operating System: A compatible operating system like Windows, Linux, or MacOS

iv. Processor: Intel Core i7

3.2.2 Software Requirements:

The software used to built the project are as listed below :

i. Python is installed

ii. Required Libraries:

a) Streamlit

b) Tensorflow

c) Pandas

d) Numpy

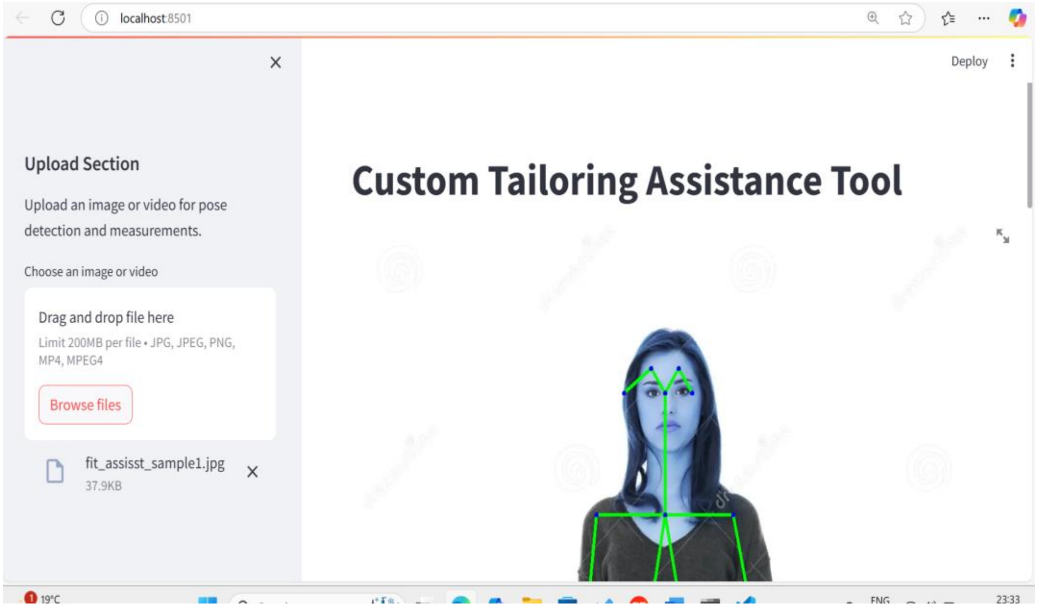
e) Open-CV

f) Media Pipe

CHAPTER 4

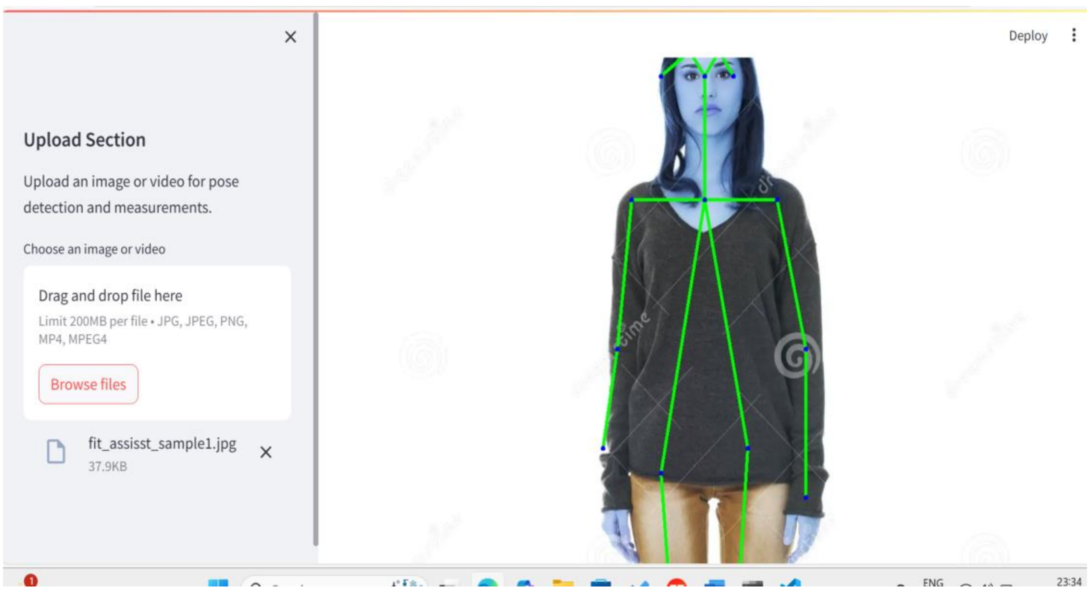
Implementation and Result

4.1 Snap Shots of Result:

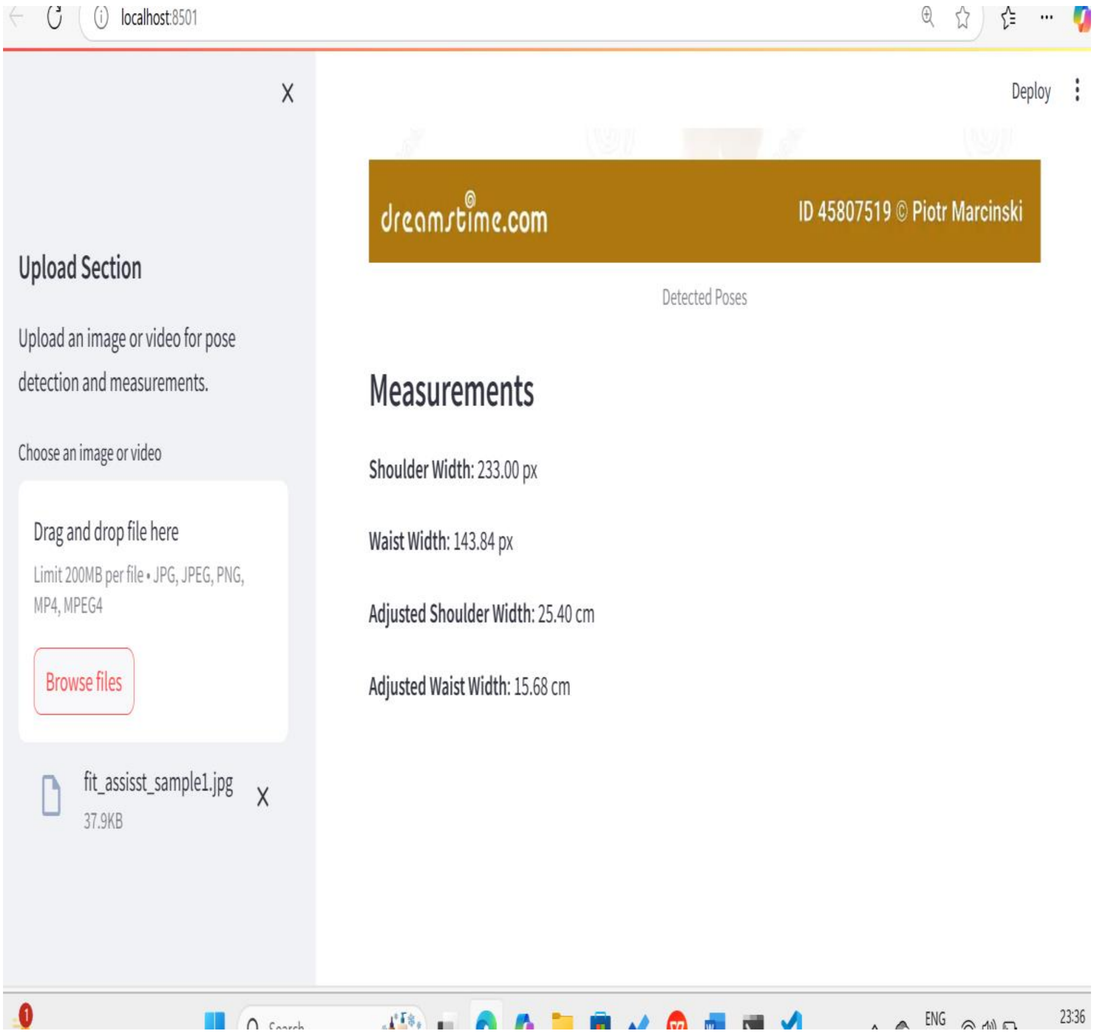


Import an image:

For body measurement using human pose estimation

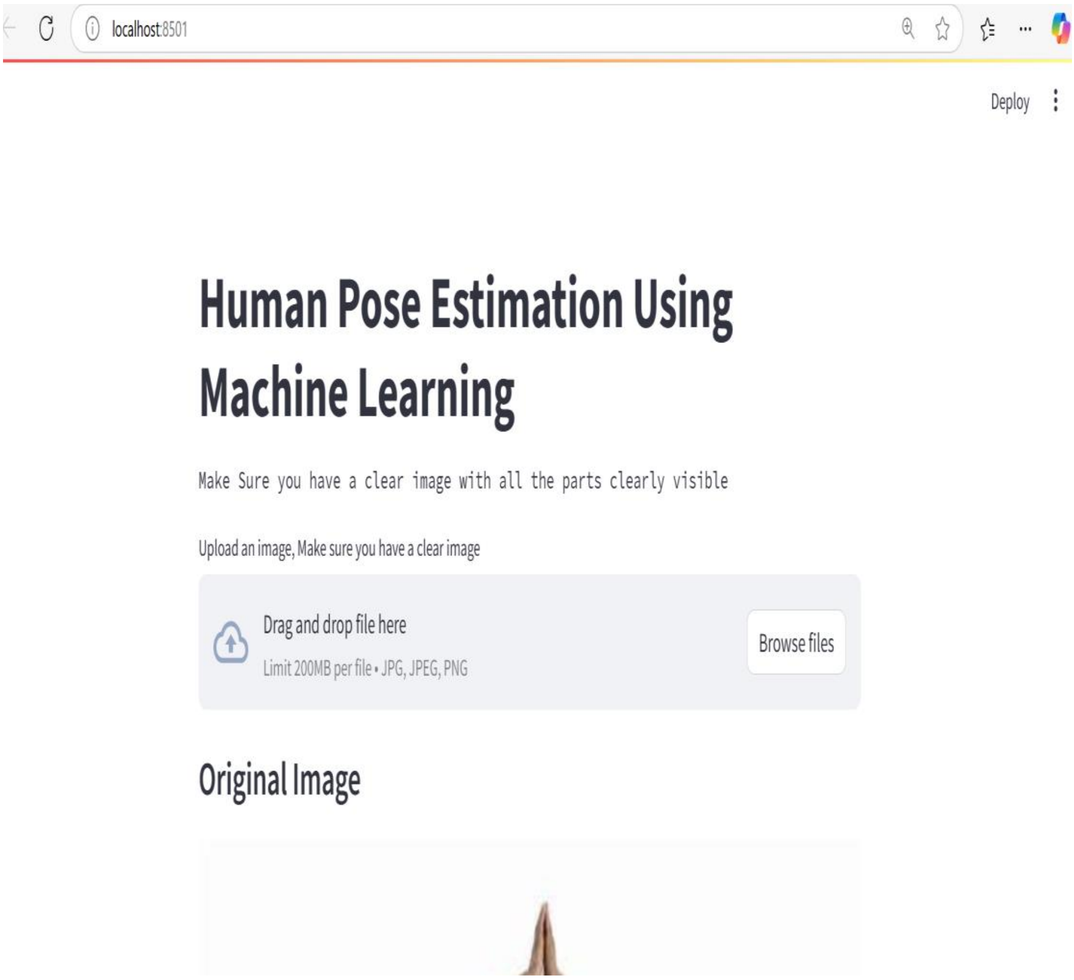


Human Pose Detection: For Body Measurement

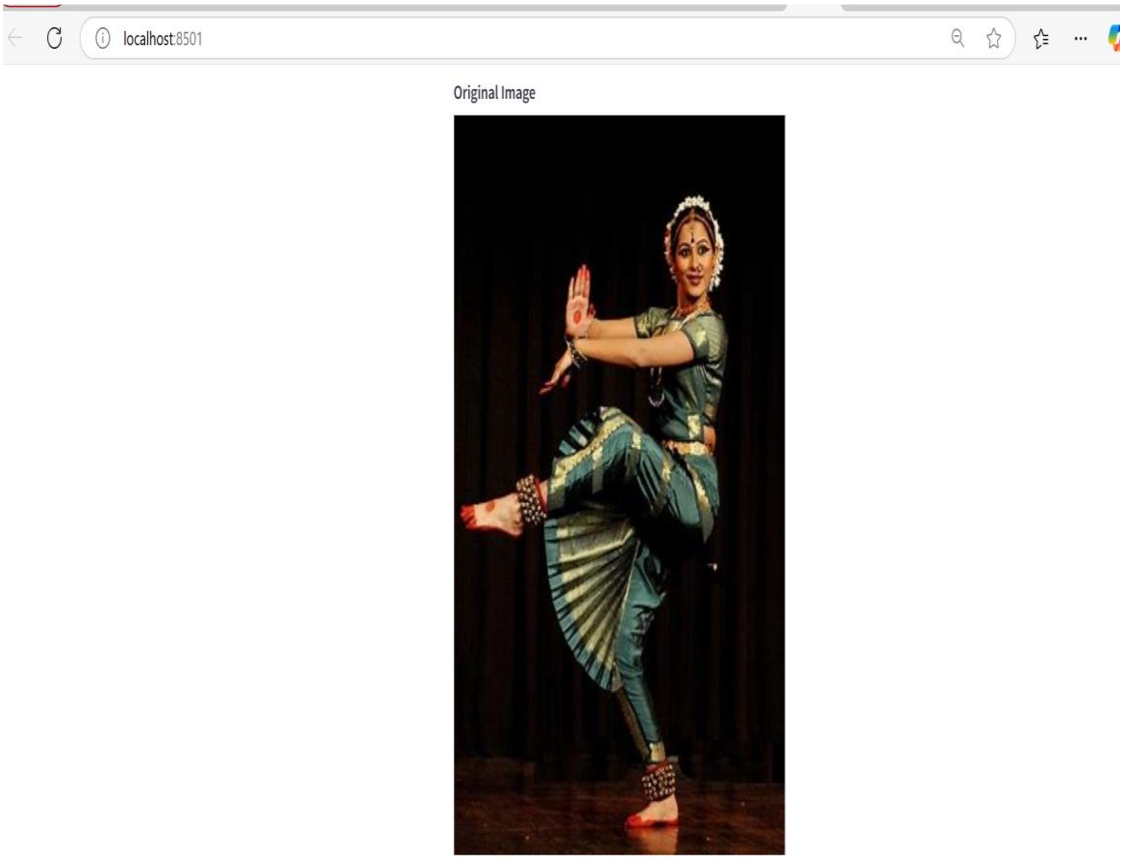

localhost:8501
Deploy
Upload Section
Upload an image or video for pose detection and measurements.
Choose an image or video
Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG, MP4, MPEG4
Browse files
fit_assisst_sample1.jpg
37.9KB
dreamstime.com
ID 45807519 © Piotr Marcinski
Detected Poses
Measurements
Shoulder Width: 233.00 px
Waist Width: 143.84 px
Adjusted Shoulder Width: 25.40 cm
Adjusted Waist Width: 15.68 cm
ENG
23:36

Estimating body measurement

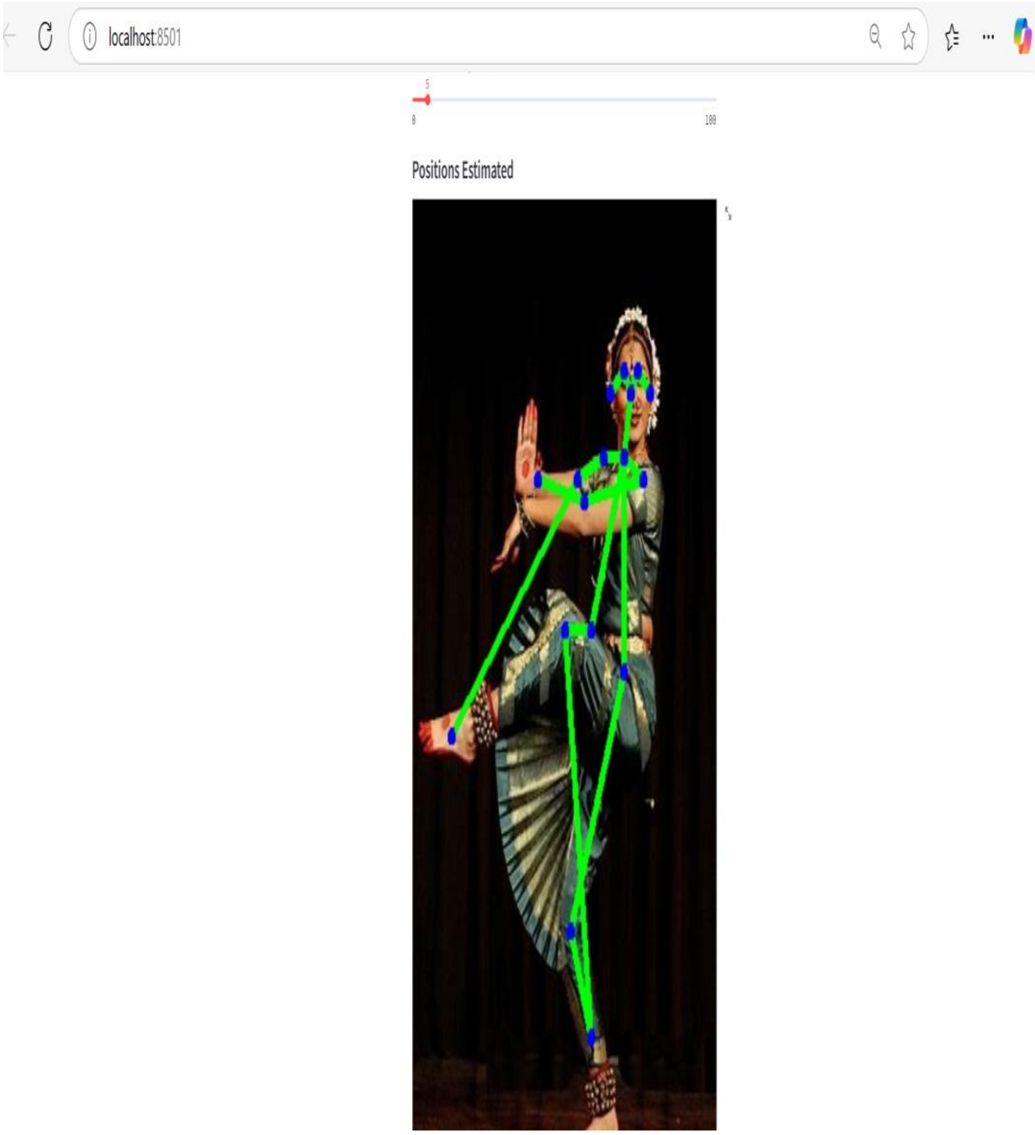
Human Pose Estimation Output:



The image selector: For Human Pose Estimation



The image selected: For Human Pose Estimation



With threshold set, pose is estimated

4.2 GitHub Link for Code:

https://github.com/Learner015/Microsoft_SAP_Edunet_Internship_Final_Project.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Human pose estimation for virtual body measurement, while advanced, still grapples with achieving centimeter-level accuracy. Future research in this field will focus on enhancing the precision of body measurements. This will likely involve the development of more sophisticated algorithms that can better handle variations in body shapes, poses, and clothing. Integrating high-resolution imaging and advanced machine learning models will be critical to refining measurement accuracy. Additionally, leveraging multi-view systems and incorporating depth sensors could provide more detailed and accurate body measurements. Ethical considerations, such as user privacy and data security, will also need to be addressed as the technology evolves and becomes more widespread.

5.2 Conclusion:

Human pose estimation has the potential to revolutionize virtual body measurement, but current technology falls short of achieving the centimeter-level accuracy required for many practical applications. Continued research and development are essential to overcome these limitations. By focusing on improved algorithms, integrating advanced imaging techniques, and addressing ethical concerns, the field can move closer to achieving the desired accuracy. As these advancements are realized, human pose estimation will play a crucial role in various applications, from virtual fitting rooms to personalized health and fitness assessments, enhancing user experience and reliability.

REFERENCES

[1]. Dubey, S., Dixit, M. A comprehensive survey on human pose estimation approaches. *Multimedia Systems* **29**, 167–195 (2023). https://doi.org/10.1007/s00530-022-00980-0

[2]. Gao, Z., Chen, J., Liu, Y. *et al.* A systematic survey on human pose estimation: upstream and downstream tasks, approaches, lightweight models, and prospects. *Artif Intell Rev* **58**, 68 (2025). https://doi.org/10.1007/s10462-024-11060-2

[3]. Samkari E, Arif M, Alghamdi M, Al Ghamdi MA. Human Pose Estimation Using Deep Learning: A Systematic Literature Review. *Machine Learning and Knowledge Extraction*. 2023; 5(4):1612-1659. https://doi.org/10.3390/make5040081