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

Kanraj Dhanush, dhanushkanraj001@gmail.com

Under the Guidance of

P.Raja, Master Trainer, Edunet Foundation

ACKNOWLEDGEMENT

We would like to take this opportunity to express our heartfelt gratitude to all individuals who have supported us directly or indirectly during the course of this thesis work.

Firstly, we would like to extend our sincere thanks to **P. Raja Sir** for his invaluable guidance and unwavering support throughout this project on **Human Pose Estimation using Machine Learning**. His deep knowledge, insightful inputs, and methodical approach were instrumental in helping us grasp the intricate details of the project.

The way he patiently explained the concepts and ensured we understood every aspect of the work provided us with the confidence to execute this project successfully. His dedication to teaching and mentoring not only aided in the completion of this project but also left a lasting impact on our professional growth and learning.

We are truly privileged to have had the opportunity to work under his mentorship and are profoundly grateful for his encouragement and assistance.

ABSTRACT

This project focuses on identifying human poses using machine learning, which can be useful in many areas like yoga, running, sleep posture monitoring, and sports analysis. For example, it can analyze a bowler's head and hand positions in cricket or a football player's kicking posture when aiming for a goal. This information can provide valuable feedback to improve performance. The main problem is that existing methods for pose estimation often struggle with accuracy, handling occlusions, and working in different scenarios. The goal of this project is to create a system that can detect human poses accurately and give detailed insights that help people improve their activities. The project uses advanced machine learning techniques with tools like OpenCV and MediaPipe. The project uses advanced machine learning techniques, integrating Streamlit to build an interactive and user-friendly interface for visualizing and analyzing pose estimation results. This platform allows users to input images or videos, view the detected poses. Streamlit makes the system accessible and easy to use, ensuring that users can analyze poses without needing extensive technical knowledge. The system identifies key points on the human body, such as joints, to recreate poses and analyze movements. The results show that the system works well, even in challenging situations like complex movements or when parts of the body are hidden. It provides accurate pose detection and meaningful analysis, making it useful for sports, fitness, and everyday activities. In conclusion, this project demonstrates how machine learning can be used to identify and analyze human poses effectively. By using Streamlit for an interactive interface, the system offers a practical solution that gives useful feedback and helps people perform better in various fields like sports, health, and daily life

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

This project focuses on human pose estimation, a technology that allows a computer to detect and analyze the positions of various body parts in images or videos. By identifying key points like the head, shoulders, elbows, knees, and ankles, the system can accurately map out how the body is positioned in space. This technology has numerous applications across different fields, including sports, fitness, healthcare, and even everyday activities. The aim of this project is to develop a system that can provide real-time feedback on human poses to help improve performance and reduce the risk of injuries.

In today's world, many people engage in physical activities, whether it's in a gym, yoga center, or while playing sports. However, many individuals struggle with maintaining the correct posture or body alignment during exercises or movements, which can lead to ineffective results or even harm. This project aims to address that problem by offering a solution that not only detects human poses but also provides helpful feedback on how to improve them. Whether someone is practicing yoga, lifting weights, or participating in sports like football or cricket, accurate pose detection can enhance their performance and prevent injuries.

The system is built using machine learning techniques, particularly with tools like OpenCV and MediaPipe, which are known for their efficiency in real-time pose detection. To make the system easy for everyone to use, it's integrated with Streamlit, which provides a simple interface for users to interact with. The goal is to create a tool that is accessible, user-friendly, and capable of providing valuable insights into posture and movement. With this project, we hope to make human pose estimation technology more widely available and applicable to anyone looking to improve their physical activities, whether for fitness, health, or sports performance.

1.1Problem Statement:

Human pose estimation is a significant challenge in computer vision, involving the identification and analysis of human body postures from images or videos. Existing systems often struggle with variations in lighting, occlusions, and complex movements, limiting their reliability in real-world scenarios. This problem is crucial in fields like sports analysis, healthcare, and fitness, where accurate pose detection can provide valuable insights and feedback. Addressing these challenges through an advanced and user-friendly system can enhance performance, prevent injuries, and open opportunities for diverse applications.

1.2Motivation:

I first observed the need for a better pose estimation system while spending time at a local gym. I saw many people performing exercises without knowing if their form was correct or not. Some were doing squats with improper posture, which could lead to injuries. Others were lifting weights, but their arms and legs weren't in the right positions, which could have been avoided with better guidance. I also noticed similar challenges in a yoga center, where practitioners were struggling to align their bodies properly during poses, even though proper alignment is crucial for effective practice and injury prevention. It struck me that if we could provide real-time feedback on their body postures, it could help them correct their form on the spot. This led me to realize how beneficial a system like this could be for people in everyday activities, not just in the gym or yoga center. Whether someone is playing a sport, performing physical therapy exercises, or simply trying to improve their posture, accurate pose detection and feedback could make a big difference. That's why I chose to develop this project, to bridge this gap by creating a simple, user-friendly system that gives people the real-time feedback they need to improve their body alignment and movements.

1.3Objective:

The objectives of this project are to develop a system that accurately detects human poses from images or videos, providing valuable insights into body alignment and movements. A key goal is to create a user-friendly interface, ensuring that the system is easy to interact with, even for non-technical users. The system aims to offer real time feedback on human poses, with a specific focus on improving sports performance, fitness routines, and rehabilitation exercises. By integrating tools like Streamlit, the project strives to make pose detection and analysis accessible to a wide range of users, allowing them to receive immediate guidance on their posture and movements.

1.4Scope of the Project:

The scope of this project includes the development of a human pose estimation system that can be applied in various fields such as sports, fitness, healthcare, and wellness. The system can be used in settings like meditation centers, yoga studios, gyms, and sports arenas for activities such as football, cricket, kabaddi, running, long jump, and high jump. It provides real-time feedback on body posture, helping individuals improve their technique, prevent injuries, and enhance overall performance. The system can also be beneficial for rehabilitation exercises and monitoring posture during daily activities. However, the limitations of this project include its current focus on 2D pose estimation, which may not be as effective in dynamic or highly complex environments. The system might face challenges in handling occlusions or detecting poses in crowded areas. Additionally, the real-time feedback feature may require high computational resources for faster processing, especially when analyzing rapid movements in sports. In the future, the scope of the project could be expanded to include 3D pose estimation, improving accuracy and providing more detailed analysis. The system could also be integrated into mobile applications for on-the-go monitoring, and its use could extend to various other fields such as dance, rehabilitation for elderly people, and performance analysis in more complex sports. Enhancements in AI and machine learning could further optimize the system to work in more diverse environments and under varying conditions.

CHAPTER 2

Literature Survey

With the rapid growth of Machine Learning (ML) and Deep Learning (DL) models, it has become very important to test and evaluate them to ensure they work accurately. A good model should ideally have an accuracy of above 95% for both training and testing datasets. However, achieving this is not always easy due to challenges like not having enough high-quality data and the model being too sensitive to small changes in the input.

These challenges make it necessary to improve models so they can handle different situations and avoid errors. Testing also helps make sure the model is not overfitting (working well only on the training data) or underfitting (not learning enough from the data). As these models are used for tasks like identifying objects, tracking movements, and understanding poses, proper testing ensures they work effectively in real-world applications.

2.1About The Project

This project focuses on creating a system that can identify and analyze human poses in real time using advanced tools like OpenCV and MediaPipe. The system detects key points on the body, such as shoulders, elbows, knees, and ankles, and uses these to recreate and study the body's posture and movements. This technology is especially helpful in areas like sports, fitness, and healthcare, where understanding and improving body movements is important.

The goal of the project is to help people improve their posture and performance during different activities. For example, in sports like football or cricket, the system can analyze how players move, such as a bowler's arm position or a footballer's kick, and provide feedback to improve their skills. In fitness and yoga, the system ensures that exercises and poses are done correctly to avoid injuries and make workouts more effective. The system gives real-time feedback, allowing users to fix mistakes immediately, which makes it useful for both beginners and professionals.

To make it easy for everyone to use, the system is built with Streamlit, a platform that provides a simple and interactive interface. Users can upload images or videos and quickly

see the analysis of their poses. In the future, the system can be improved to provide 3D pose detection for even more accurate analysis. This will open up new possibilities in areas like sports coaching, rehabilitation, and fitness training. By combining advanced technology with user-friendly tools, this project makes it easier for people to monitor and improve their physical activities, leading to better results in sports, health, and fitness.

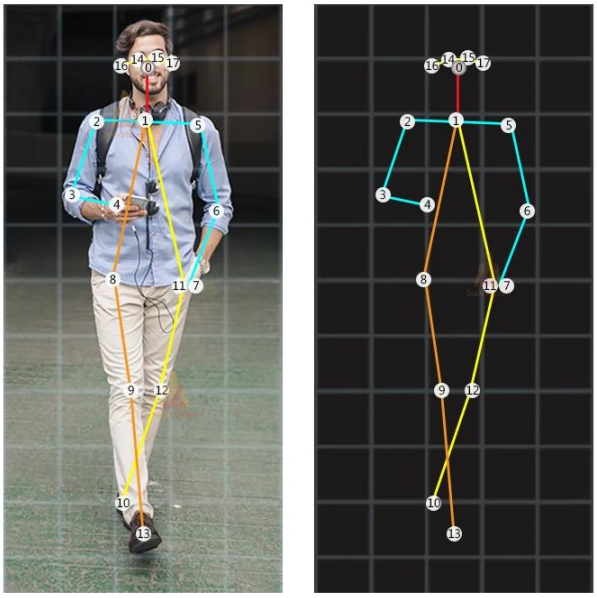

Fig1:humanPoseDetection

2.2Literature Review

Deep Learning-based Human Pose Estimation [1]: This paper presents Human pose estimation aims to locate the human body parts and build human body representation (e.g., body skeleton) from input data such as images and videos.

Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., ... & Shah, M. (2023). Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, *56*(1), 1-37.

Human pose estimation and its application to action recognition [2]: This paper presents a estimation aims at predicting the poses of human body parts in images or videos. Since pose motions are often driven by some specific human actions, knowing the body pose of a human is critical for action recognition.

Song, L., Yu, G., Yuan, J., & Liu, Z. (2021). Human pose estimation and its application to action recognition: A survey. *Journal of Visual Communication and Image Representation*, *76*, 103055.

ViTPose [3]: This paper presents a domain knowledge is considered in the design, plain vision transformers have shown excellent performance in visual recognition tasks.

Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. Advances in Neural Information Processing Systems, 35, 38571-38584.

JoyPose [4]: This paper presents a Video-based 3D human pose estimation is an important yet challenging task for many human-involved pattern recognition systems.

Du, S., Yuan, Z., Lai, P., & Ikenaga, T. (2024). JoyPose: Jointly learning evolutionary data augmentation and anatomy-aware global–local representation for 3D human pose estimation. Pattern Recognition, 147, 110116.

CHAPTER 3

Proposed Methodology

3.1 System Design

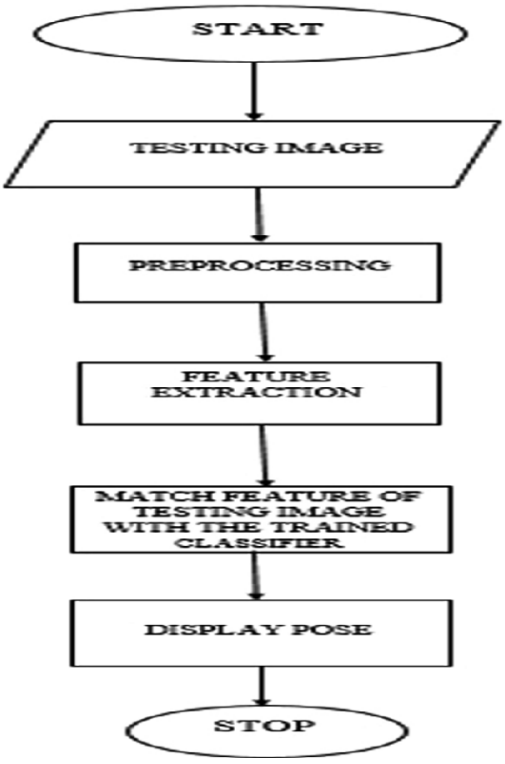

Fig2: AMachineLearningApproach

The system is designed to detect human body poses from images or videos. The diagram below illustrates the workflow:

- **Input**: Image or video of a person.
- **Pose Detection**: The system uses pre-trained models to detect key points on the human body.
- **Analysis**: The system analyzes the detected poses, providing feedback on areas like posture and movement.

- **Output**: Real-time feedback and analysis shown to the user via an interactive interface (Streamlit).

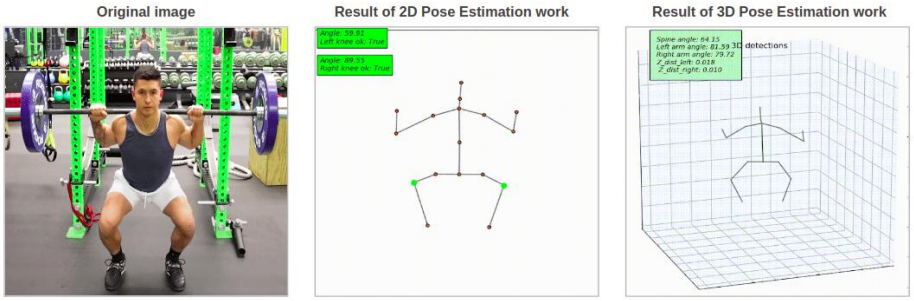

Fig3 : Thedifferencebetween2Dand3DposeEstimationReconstructions

3.2 Requirement Specification

3.2.1 Hardware Requirements:

- A computer with a minimum of 4GB RAM.
- A webcam (if using real-time video input).
- GPU (optional, for faster processing).

3.2.2 Software Requirements:

- opencv_python_headless==4.5.1.48
- streamlit==0.76.0
- numpy==1.18.5
- matplotlib==3.3.2
- Pillow==8.1.2

CHAPTER 4

Implementation and Result

4.1 Result:

In a Human Pose Estimation project, the primary objective is to identify and visualize the positions of key body parts, such as the head, shoulders, elbows, wrists, hips, knees, and ankles, within an image or video. The process involves detecting these body landmarks and connecting them with lines to form a skeleton-like representation, highlighting the relationships between joints and limbs. This allows for a detailed understanding of human body movements in a static or dynamic context. Pose estimation is valuable in various fields, including sports analysis, fitness tracking, healthcare diagnostics, animation, and virtual reality. By applying advanced machine learning techniques, such as convolutional neural networks (CNNs), the model can process complex visual inputs and accurately predict the pose in real-time. The results are often presented through annotated images or videos, which can be displayed on a screen or saved for further analysis. This project enhances the ability to track, monitor, and optimize human motion.

Fig4: StandingInput1

The input image is passed through the pose estimation model to predict the body parts and their positions. Here's an example input image

Fig5: StandingOutput1

After completing the pose detection process, the model produces an output in the form of a fully annotated image. In this image, keypoints representing essential body parts such as the head, shoulders, elbows, wrists, hips, knees, and ankles are marked. These points are connected with lines to illustrate the human pose, effectively creating a skeletal structure of the body. This visualization provides clear insights into the alignment and movement of the individual. The annotated image not only highlights the spatial relationships between joints and limbs but also allows for further analysis of body posture and motion. This annotated output can be displayed on the screen for immediate review or saved for later use, making it suitable for applications in sports analysis, healthcare, fitness tracking, and other fields that require detailed human movement analysis. The model's accuracy in detecting and visualizing these keypoints is crucial for a wide range of practical uses.

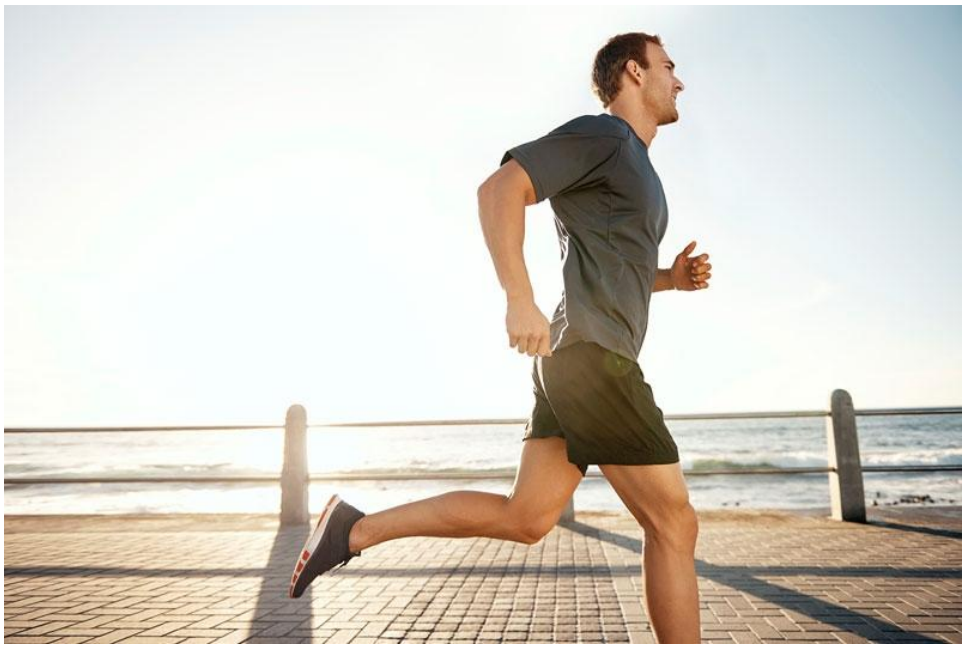

Fig6: RunningInput2

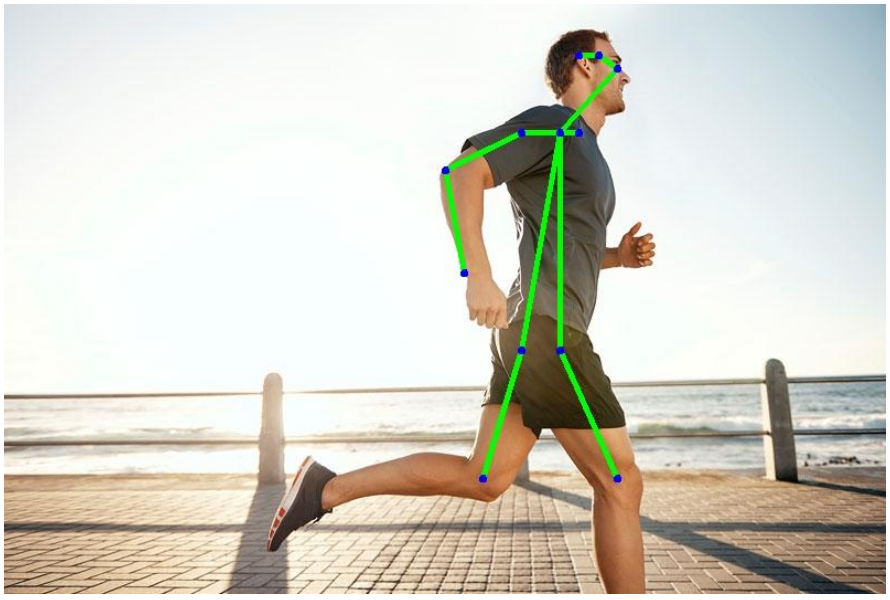

Fig7: RunningOutput2

The model has successfully completed pose detection, generating an image with keypoints and lines that represent the human pose. The resulting image can now be either displayed on the screen or saved.

4.2 GitHub Link for Code:

https://github.com/dhanushkanraj/AICTE-internship.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

To improve the system's real-time performance, it is essential to optimize the underlying algorithms, ensuring faster inference times and more efficient processing. This could involve leveraging techniques such as model pruning, quantization, or using specialized hardware like GPUs or TPUs for faster computations, enabling the system to handle more dynamic activities, such as rapid movements in sports or rehabilitation exercises. Extending the model to incorporate 3D pose estimation would further enhance the system's capabilities by capturing depth information, offering a more accurate representation of human body posture in three-dimensional space. This extension is particularly valuable in sports and rehabilitation, where understanding the angles and alignment of body parts from different perspectives is crucial for performance optimization and injury prevention. Additionally, the feedback mechanism can be significantly improved by integrating more advanced machine learning models that analyze detected poses in real-time and generate personalized suggestions for posture correction. These suggestions could include specific joint angles, alignment improvements, or movement adjustments tailored to the individual's needs. By providing more precise, actionable insights, users can receive better guidance for improving their form during exercises, sports activities, or physical therapy. Ultimately, these improvements would enhance the system's practical applications in diverse fields such as sports training, physical therapy, and fitness, making it a more robust tool for performance monitoring and injury prevention.

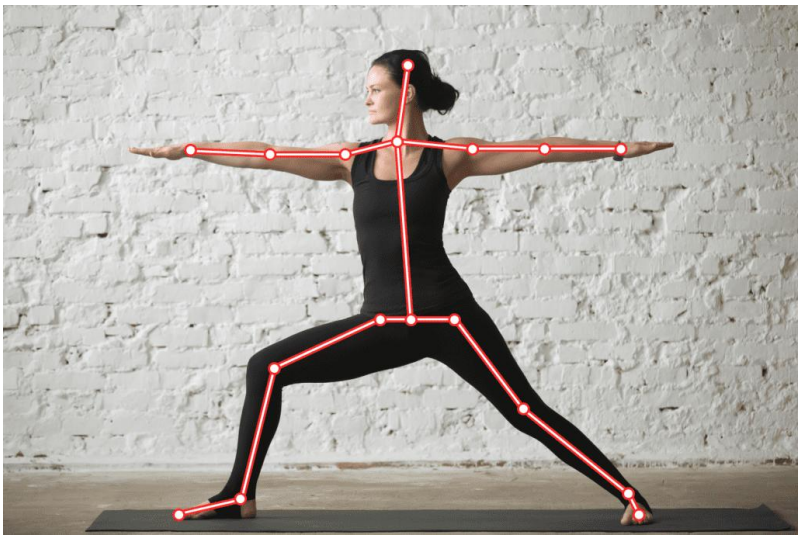

Fig8: EnablingHealthAndFitnessAppsWithHumanPoseDetection

5.2 Conclusion:

This project effectively showcases the power of machine learning in achieving accurate human pose estimation, offering a reliable and real-time solution for tracking body movements. By leveraging advanced tools like OpenCV, MediaPipe, and Streamlit, the system is able to detect and visualize key body points with remarkable precision. OpenCV facilitates efficient image and video processing, MediaPipe enhances pose detection with its pre-trained models, and Streamlit provides an intuitive interface for real-time user interaction. The integration of these tools creates a seamless and user-friendly experience, allowing users to easily visualize their posture and receive immediate feedback. The system has a significant impact across various fields, including sports, fitness, and healthcare. In sports, it aids athletes by analyzing their movements and providing insights for improving performance and technique. In fitness, it serves as a valuable tool for individuals seeking to track their exercise form, while in healthcare, it can assist in rehabilitation by monitoring posture and movement recovery. Moreover, the system's ability to detect incorrect posture and suggest corrective actions helps prevent injuries, making it a crucial tool for both professionals and everyday users. Overall, this project represents a comprehensive and accessible solution with wide-reaching applications for enhancing human performance and well-being.

REFERENCES

[1]. Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., ... & Shah, M. (2023). Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, *56*(1), 1-37.

[2]. Song, L., Yu, G., Yuan, J., & Liu, Z. (2021). Human pose estimation and its application to action recognition: A survey. Journal of Visual Communication and Image Representation, 76, 103055.

[3]. Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. Advances in Neural Information Processing Systems, 35, 38571-38584.

[4]. Du, S., Yuan, Z., Lai, P., & Ikenaga, T. (2024). JoyPose: Jointly learning evolutionary data augmentation and anatomy-aware global–local representation for 3D human pose estimation. Pattern Recognition, 147, 110116.