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

Harsh Vardhan

harshvardhanbharti2020@gmail.com

Under the Guidance of

P.Raja, Master Trainer,Edunet Foundation

ACKNOWLEDGEMENT

I would like to extend our heartfelt gratitude to everyone who supported and guided me, both directly and indirectly, during the course of this thesis work.

First and foremost, I am deeply thankful to our supervisors, **P. Raja and Pawan Sir**, for their exceptional mentorship, steadfast encouragement, and insightful advice throughout this journey. Their guidance has been pivotal in the development and successful completion of this project, inspiring us with fresh ideas, critical perspectives, and constant motivation.

Their faith in my abilities has been a significant driving force, encouraging me to push our boundaries and strive for excellence. It has been a privilege to work under their supervision over the past weeks, and their support has enriched not only this project but also my overall academic and personal growth.

The knowledge and lessons imparted by them extend far beyond this project, shaping me into more capable professionals and responsible individuals.

I also wish to acknowledge the **TechSaksham initiative** by **Microsoft & SAP** for providing this transformative learning opportunity. Lastly, I thank my family, peers, and friends for their unwavering support and encouragement.

ABSTRACT

Human body motion estimation is a crucial task in computer vision, with applications spanning healthcare, sports, security, and robotics. This project aims to develop a system that can detect and analyze human body movements in real-time, focusing on understanding human motion patterns using advanced algorithms and models.

The project utilizes well-known machine learning and computer vision tools to achieve this goal. **Opencv** and **MediaPipe** models are used to detect key points on the human body, creating a skeleton representation of the movements. OpenCV is used for video processing, enabling real-time performance, while TensorFlow helps in pose analysis and interpretation. The system is deployed as a web-based application using Streamlit, enabling users to upload a image and generate its skeletal structure.

Key results show that the system can estimate body poses with high accuracy and process them in real-time. The system performs well in challenging environments, such as when multiple people occlude each other, under varying lighting conditions, or with cluttered backgrounds. These challenges are addressed by improving the training data and fine-tuning the model parameters, allowing the system to maintain robustness across different scenarios. The system is capable of motion tracking, gesture recognition, and activity detection, with applications in fields such as sports analysis, rehabilitation, and interactive systems.

In conclusion, this project demonstrates an effective approach to human body motion estimation, with strong potential for real-world applications in rehabilitation, sports, and human-computer interaction. The system shows promise in providing reliable real-time tracking of human motion and can serve as a foundation for further research. Future work will focus on enhancing the system's performance in diverse environments and exploring the potential of 3D motion capture to capture even more detailed body movement information.

TABLE OF CONTENT

I

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

Human pose estimation, the task of detecting and analyzing the positions and orientations of human body joints, is a fundamental challenge in computer vision with wide-ranging applications. Accurately capturing human poses in real time is essential for understanding activities such as walking, running, exercising, and even complex movements like dancing. However, achieving precise and reliable pose estimation is challenging due to variations in body shapes, occlusions, complex backgrounds, lighting conditions, and dynamic environments. Traditional methods often struggle with these challenges, necessitating advanced solutions to improve accuracy and robustness.

Figure 1: The image highlights human body motion estimation applications in sports, healthcare, HCI, security, and robotics.

This problem is significant because human pose estimation has critical applications in fields such as healthcare, sports, human-computer interaction, security, and robotics. For example, it enables efficient monitoring and analysis in physical therapy, enhances athletic performance through movement optimization, and supports gesture-based interfaces for immersive VR/AR experiences. Additionally, it aids in detecting suspicious activities for security and enables robots to interact naturally with humans.

Addressing this problem is key to driving innovation across these domains and creating intelligent systems capable of understanding and responding to human motion.

1.2 Motivation:

This project was chosen due to the growing importance of human body motion estimation in addressing real-world challenges across multiple domains such as healthcare, sports, security, and human-computer interaction. Understanding and analyzing human motion is a complex yet impactful task that enables real-time tracking, pose detection, and gesture recognition. With advancements in artificial intelligence (AI) and computer vision technologies, the project leverages powerful tools like MediaPipe and OpenCV to develop accurate and efficient solutions for this critical problem.

The potential applications of this project are extensive. It can be utilized in healthcare for rehabilitation monitoring, elderly care, and post-stroke recovery, as well as in sports for technique analysis and injury prevention. In addition, it supports gesture-based interaction in VR/AR systems, enhances intelligent surveillance for detecting suspicious activities, and aids robotics in developing assistive systems capable of natural interactions with humans.

The impact of this project is significant, as it improves healthcare monitoring, enhances sports performance, strengthens security measures, and provides intuitive user experiences through gesture recognition. By offering scalable and real-time solutions, this project lays the groundwork for advancements in intelligent systems and automation, ultimately contributing to a safer and more efficient digital and physical environment.

1.3 Objective:

The primary goal of this project is to develop a system for **real-time human body motion tracking** using **computer vision** and **machine learning** techniques. This system aims to detect and analyze human movements by estimating body poses accurately in real time.

Key objectives include:

1. **Real-Time Motion Tracking**: To create a system that captures and tracks human body movements continuously using video input, enabling real-time motion estimation.
2. **Utilizing Pose Detection Libraries**: Implement libraries like **MediaPipe** or **OpenCV** to detect key body points and estimate the pose of individuals accurately, with **OpenCV** for video processing.
3. **Improving Accuracy**: Ensure that the system functions under varying conditions (e.g., changes in lighting, occlusion) to maintain high accuracy in detecting body poses.
4. **Motion and Gesture Recognition**: Enable the system to recognize common human activities such as walking, sitting, or running by analyzing the body posture and movement patterns.
5. **Exploring Practical Applications**: Investigate the use of the system in **healthcare** for rehabilitation, in **sports** for performance analysis and injury prevention, and in **robotics** for assistive tasks.

6. **Foundation for Future Work**: Lay the groundwork for future developments in **3D motion capture**, advanced gesture recognition, and expanded applications in areas like virtual reality, gaming, and security.

These objectives aim to create an efficient, accurate, and versatile system for human motion estimation, with practical applications across several domain.

1.4 Scope of the Project:

The scope of this project encompasses the development and implementation of a human body motion estimation system that leverages computer vision and machine learning techniques to accurately detect, track, and analyze human body poses and movements in real time.

Scope of the Project:

1. **Real-Time Human Motion Estimation**: The project focuses on developing a system for detecting and tracking human body movements in real-time using computer vision techniques. It includes pose estimation, motion analysis, and gesture recognition.
2. **Use of Existing Libraries**: The project leverages popular libraries like **MediaPipe** and **OpenCV** for pose detection and **OpenCV** for video processing. These tools provide a reliable and efficient framework for implementing motion estimation.
3. **Applications in Healthcare, Sports, and Robotics**: The system can be used in various domains, such as:
    a. **Healthcare**: For rehabilitation, monitoring patients' movements, and aiding in physical therapy.
    b. **Sports**: For performance analysis, injury prevention, and technique improvement.
    c. **Robotics**: For developing assistive robots that interact with humans based on detected motion.
4. **Gesture and Activity Recognition**: The project also aims to implement simple gesture recognition algorithms to classify different human movements like walking, sitting, or running.
5. **Future Development**: This system serves as a foundation for future research, such as extending the system to 3D motion capture and integrating advanced gesture recognition techniques.

Limitations of the Project:

1. **Single-Person Limitation**: The system is primarily designed for **single-person motion tracking**. Handling multiple people in the same frame with high accuracy can be more challenging and would require more complex solutions (e.g., multi-person pose detection).
2. **Accuracy in Complex Environments**: The system may face challenges in environments with poor lighting, extreme occlusion, or highly dynamic backgrounds, leading to reduced accuracy in pose detection and motion estimation.

3. **Processing Power**: Real-time video processing and pose estimation require significant computational resources. The performance of the system may vary based on the hardware being used, particularly in resource-constrained environments.
4. **Limited Gesture Recognition**: While the system can recognize basic movements, more complex or subtle gestures may require further refinement in the gesture recognition algorithms.
5. **Dependence on Quality of Input**: The quality of the input video (resolution, frame rate, etc.) directly affects the accuracy and efficiency of the motion estimation system.In conclusion, the project provides a functional system for human motion estimation with wide-ranging applications but has limitations in terms of handling multiple people, accuracy in complex environments, and system resource requirements.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

Human body motion estimation has been a prominent area of research, with various studies focusing on overcoming challenges like pose variation, occlusion, lighting, and dynamic environments. Traditional approaches, such as edge detection and template matching, as reviewed by Yang et al. (2002) in *"Detecting Faces in Images: A Survey"*, laid foundational insights into handling visual challenges. However, these methods lacked robustness for real-world scenarios, limiting their application to pose estimation tasks.

The introduction of deep learning frameworks revolutionized the field. For example, Toshev and Szegedy (2014) proposed *DeepPose*, the first approach to use convolutional neural networks (CNNs) for directly predicting joint positions. This method significantly improved accuracy and efficiency, establishing a new standard in pose estimation. However, it struggled with occlusion and lacked temporal consistency. Similarly, Ouyang et al. (2014) enhanced pose estimation accuracy by integrating body part detection and spatial relationships through a multi-source deep learning approach, which improved robustness in complex scenarios. Despite these advancements, the high computational demands limited their real-time applicability.

Recent advancements have focused on addressing these limitations. Grinciunaite et al. (2016) explored 3D convolutional neural networks (3D CNNs) to incorporate spatial and temporal information, ensuring smoother motion tracking. While this approach addressed temporal consistency, it faced challenges with computational efficiency. Practical tools like OpenPose and MediaPipe, demonstrated by Code Arts (YouTube), have made real-time applications more accessible, though they do not yet fully explore advanced methods like 3D pose estimation or temporal modeling. Moreover, Gupta et al. (NIT Kurukshetra) highlighted the utility of pose data for real-world tasks like healthcare and sports, using machine learning algorithms such as SVM and Random Forest for activity recognition. However, their approach lacked temporal modeling, limiting its effectiveness in continuous motion analysis.

These studies highlight both the progress and challenges in human body motion estimation, emphasizing the need for robust, real-time solutions that integrate spatial and temporal data for seamless motion tracking in dynamic environments.

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

1. DeepPose (Alexander Toshev & Christian Szegedy, 2014)

DeepPose was the first framework to leverage deep learning for human pose estimation by using Convolutional Neural Networks (CNNs) to predict joint coordinates as a regression problem. It replaced traditional handcrafted feature extraction methods with end-to-end learning, improving pose estimation accuracy. However, it struggled with frame-to-frame inconsistencies and lacked the temporal context necessary for smooth motion tracking..

2. OpenPose

OpenPose is a popular open-source tool for real-time multi-person pose detection using a bottom-up approach. It combines CNNs with a part affinity field to identify body parts and their relationships. OpenPose is highly efficient in multi-person pose estimation and is widely used in sports and healthcare applications. However, it focuses on 2D pose estimation and has limitations in temporal dynamics and 3D pose analysis.3.

3. 3D CNNs (Agne Grinciunaite et al., 2016):

3D Convolutional Neural Networks (3D CNNs) process sequences of images to capture both spatial and temporal dynamics, which helps in smooth pose estimation and motion tracking. This method addresses the need for temporal consistency in human motion analysis. However, the high computational requirements of 3D CNNs limit their real-time application and adaptability to diverse environments.

4. MediaPipe Pose

MediaPipe Pose is a real-time pose detection framework developed by Google, which uses machine learning to predict 33 human body landmarks in both 2D and 3D. Optimized for real-time applications, MediaPipe features a lightweight architecture with multi-platform support. However, it faces challenges in accurately detecting complex poses and occluded body parts.

5.Multi-Source Deep Learning (Wanli Ouyang et al., 2014):

This approach combines body part detection, spatial relationships, and contextual information into a unified framework for improved pose estimation. By leveraging contextual data, it enhances robustness in scenarios involving occlusions and complex body configurations. However, it is computationally expensive and primarily designed for static pose estimation, limiting its use for dynamic or continuous motion analysis.

6. **Machine Learning Algorithms for Activity Recognition (Abhay Gupta et al.):**

This approach integrates pose estimation with classifiers like Support Vector Machines (SVM) or Random Forests to recognize human activities based on skeletal features. It demonstrates the utility of pose data for applications like human activity recognition (HAR) in healthcare and sports. However, it struggles with noisy or incomplete data and lacks temporal modeling for continuous motion analysis.

2.3 Gaps or limitations in existing solutions

Despite significant progress in human pose estimation, several gaps and limitations persist:

1. **Limited Dataset Variability:** Many models are trained on narrow datasets, which limits their ability to generalize to new environments, varied poses, or different body types, impacting their real-world applicability.

2. **Lack of Temporal Consistency:** Many pose estimation models focus on individual frames, struggling to maintain consistency across video sequences, leading to unstable predictions in motion tracking.

3. **High Computational Costs:** Advanced models like 3D CNNs or multi-source deep learning techniques require substantial computational resources, limiting their use in real-time applications or on devices with constrained processing power.

4. **Occlusion Handling:** Current systems often struggle with occlusions, where parts of the body are hidden by objects or other people, affecting pose detection accuracy.

5. **Limited 3D Pose Estimation:** Most models, such as OpenPose, focus on 2D pose estimation, missing critical depth information for more accurate motion tracking and activity analysis in dynamic or complex environments.

CHAPTER 3

Proposed Methodology

3.1 System Design:

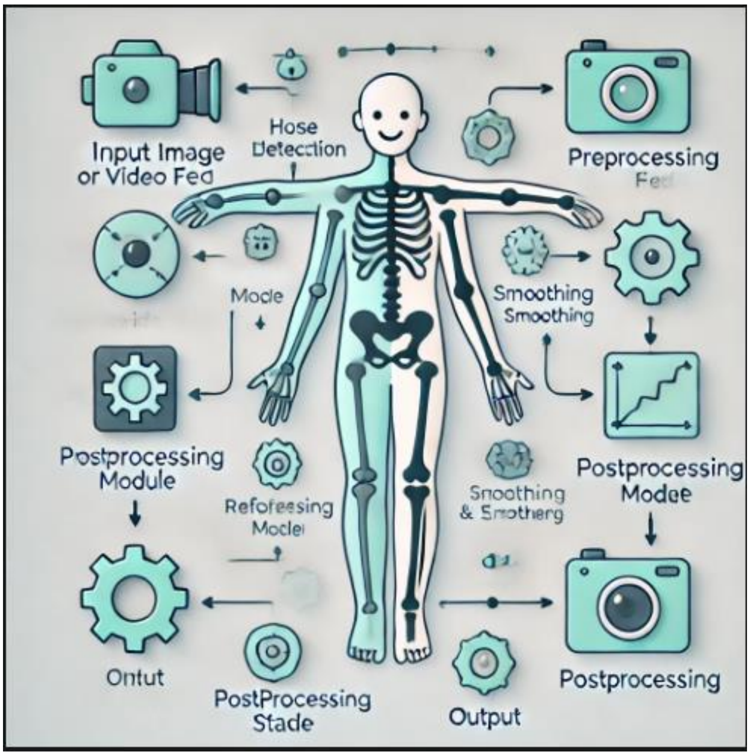

Figure 2: Human pose estimation: From raw image data to refined skeletal tracking. A journey through preprocessing, smoothing, and post-processing for accurate and insightful motion analysis

The Human pose estimation system involves several key components working together. Here's the breakdown of the system architecture:

1.Data Collection: The system starts by collecting real-time video streams or static images as input, either through a webcam or from a predefined dataset.

2.Data Preprocessing: The collected frames are pre-processed to resize and normalize the images, ensuring they are in a format suitable for pose detection. This step also includes enhancing the quality of frames if necessary.

3.Pose Detection: MediaPipe's pose detection model is applied to the pre-processed frames to detect key human body landmarks (33 points, such as joints and limbs) in real-time. This is the core step where pose estimation takes place.

4.Post-Processing: After detecting the landmarks, the system applies additional processing techniques like filtering and smoothing to ensure stable and consistent pose tracking over time.

5.Activity Recognition (Optional): If the system is designed to recognize human activities (e.g., walking, running), TensorFlow is used to train a model that classifies activities based on the detected pose landmarks.

6.Visualization: The detected body landmarks are overlaid onto the original frames, providing a visual representation of the detected poses in real-time. This allows users to see the motion tracking in action.

7.Real-Time Output: Once the pose detection and optional activity recognition models are trained, the system can classify and visualize human poses and activities from new video streams or images in real-time, providing instant feedback to the user.

This system efficiently integrates MediaPipe for pose estimation, TensorFlow for activity recognition (if needed), and OpenCV for video handling, providing a seamless and real-time solution for human pose detection and analysis.

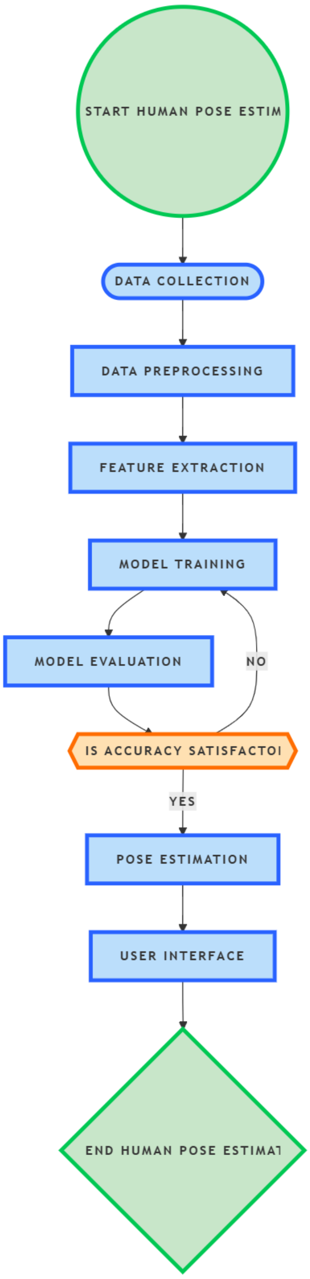

Figure 3: Workflow

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

1. **Processing Unit**:
   a. **CPU**: Dual-core or quad-core processor (e.g., Intel i5 or Ryzen 3/5).

b. **GPU**: If unavailable, rely on integrated graphics (e.g., Intel UHD Graphics) or offload intensive tasks to cloud platforms like Google Colab for free GPU access.

2. **Camera**:
   a. Basic 720p or 1080p webcam (e.g., Logitech C270) for capturing video frames.
   b. Alternatively, use a smartphone camera with tools like DroidCam for video input.

3. **Memory and Storage**:
   a. **RAM**: Minimum of 8GB for efficient operations.
   b. **Storage**: At least 256GB HDD or SSD to store datasets and lightweight models.

3.2.2 Software Requirements:

- **Programming Languages**:
  Python: Core language for development due to its rich ecosystem for machine learning and image processing.

- **Deep Learning Frameworks**:
  MediaPipe: Lightweight and highly optimized for real-time pose estimation.
  TensorFlow Lite: For optimized models on standard systems.
  Visualization Tools:
  OpenCV: For image processing and pose overlay visualization.
  Data Processing and Augmentation:
  NumPy and **Pandas**: For basic data handling and feature extraction.
  Albumentations: For augmenting input frames for robustness.

- **Pre-trained Models**:
  PoseNet (MobileNet-based): Efficient and accurate for pose estimation.
  BlazePose: Optimized for real-time applications with low computational requirements.

- **Development Environment**:
  Google Colab: Free access to cloud GPUs for training and testing models.
  Jupyter Notebook or **VS Code**: For local development.

- **Operating System**:
  Windows, **Linux**, or **macOS** – all compatible with Python-based libraries.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Kindly provide 2-3 Snapshots which showcase the results and output of your project and after keeping each snap explain the snapshot that what it is representing.

[1] First Snapshot:

Figure 4: A person in a dynamic pose.

Figure 5: Pose estimation with key points and skeletal connections.

The snapshot represents human pose estimation, a technique in computer vision where the goal is to detect and map the key points of a human body from an image or video. Here's an explanation of each side of the image:

1. **Real Image**:
    a. This is the original image of a person posing in a dynamic and expressive posture.
    b. No additional processing or overlays are present.
2. **Output Image**:
    a. This is the result of applying a human pose estimation model.
    b. Keypoints corresponding to significant parts of the human body (e.g., head, shoulders, elbows, wrists, hips, knees, and ankles) are marked with green dots.

c. These keypoints are connected with red lines to form a skeletal representation of the person's pose.
d. The model has successfully identified the position and orientation of the person based on their body structure.

[2] Second Snapshot:

Figure 6: A person in a dynamic pose.

Figure 7: Pose estimation applied to the T-pose with key points and skeletal structure highlighted.

The snapshot depicts another example of human pose estimation in action, where a model identifies and maps the key points of a human body based on a static pose. Here's an explanation of each side of the image:

1. **Real Image**:
   a. This is the original image of a person standing in a "T-pose" with arms extended horizontally to the sides.
   b. No additional processing or annotations are applied.
2. **Output Image**:
   a. This image demonstrates the output of a pose estimation model.
   b. Green dots represent the detected key points of the body, such as the nose, eyes, shoulders, elbows, wrists, hips, knees, and ankles.
   c. Red lines connect these key points to form a skeletal representation of the person's pose.
   d. The model has effectively captured the symmetry and alignment of the pose, clearly showing the extended arms and straight posture.

[3] Third Snapshot:

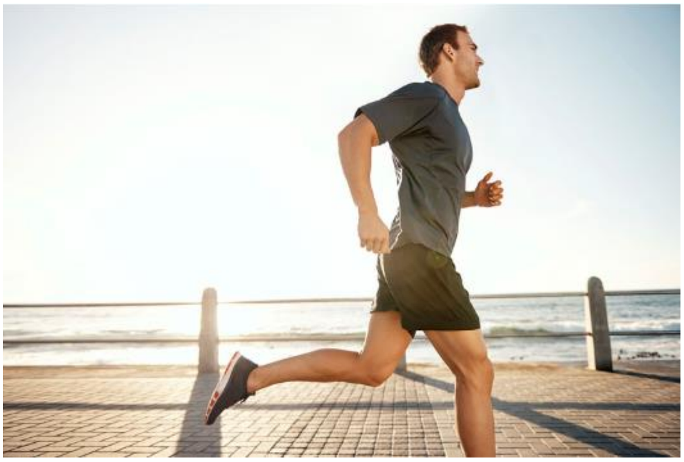

Figure 8: A person running on a boardwalk by sea

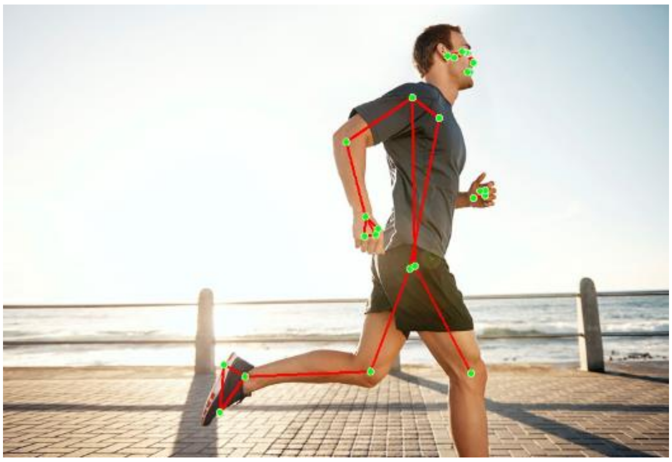

Figure 9: Pose estimation applied to the running motion, highlighting key joints and skeletal alignment.

This snapshot represents human pose estimation applied to a dynamic activity—running. Here's the breakdown of the image:

1. **Real Image**:
   a. The original image shows a person in motion, running outdoors near a railing and a body of water.
   b. No processing or annotations are visible; this serves as the input for the pose estimation model.
2. **Output Image**:
   a. This is the processed output of the pose estimation model.
   b. Green dots mark the key points of the body, including facial features, shoulders, elbows, wrists, hips, knees, and ankles.
   c. Red lines connect these key points, forming a skeletal overlay that illustrates the posture and movement of the runner.
   d. The model captures the bending of the arms and legs, showing the running stride and overall posture.

4.2 GitHub Link for Code:

https://github.com/Harsh04-Vard/Human-Pose-Estimation

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Provide suggestions for improving the model or addressing any unresolved issues in future work.

To further enhance the Human Pose Estimation model and address any unresolved challenges, the following suggestions could be considered for future work:

1. Real-Time Pose Estimation:

- Currently, the app processes static images, but adding real-time video feed processing (e.g., from a webcam or mobile camera) would make it more dynamic and applicable to real-time applications, such as live fitness tracking or virtual training sessions.
- Implementing efficient real-time video processing with reduced latency can be a key enhancement, especially for use cases like gaming or VR.

2. Accuracy and Model Enhancement:

- **Higher Model Complexity:** While the model complexity is set to 1 for faster processing, increasing it could improve accuracy, especially in challenging environments (e.g., occluded body parts, complex poses).
- **Custom Training:** Training the model on domain-specific datasets (e.g., sports poses or medical postures) can improve the model's accuracy and robustness for specific use cases, such as posture correction for physiotherapy.
- **Multi-Person Detection:** Currently, the model is optimized for detecting a single person. Extending it to handle multiple people in the same frame would be useful in crowd-based applications or events where multiple subjects need to be tracked simultaneously.

3. Incorporating 3D Pose Estimation:

- Moving from 2D pose estimation to 3D can significantly improve the depth perception and precision of motion tracking, especially when analyzing activities that require understanding of body depth (e.g., gym exercises or sports movements).
- 3D pose estimation would help in providing more accurate feedback on joint angles and body alignment.

4. Integration with Other AI Systems:

- **Gesture Recognition:** Integrating pose estimation with gesture recognition models can expand the app's functionality to understand user interactions, enabling applications in gaming, sign language recognition, or smart home control.
- **Activity Recognition:** Combining pose estimation with activity recognition (e.g., walking, jumping, stretching) could make the system more intuitive, allowing it to detect specific actions or movements in real-time and provide context-aware feedback.

5. User Feedback and Personalization:

- Incorporating user-specific feedback into the system could help personalize the pose estimation results. For instance, the app could track posture improvements over time and suggest exercises based on the user's performance or alignment issues.
- **Integration with Wearable Devices:** Integrating the model with wearable sensors (e.g., motion trackers, smartwatches) could provide more detailed data, offering more personalized feedback and insights about a user's movement.

6. Handling Environmental Factors:

- The current model may struggle with low-light environments, occlusions, or poor-quality images. Future improvements could focus on enhancing robustness against these factors, using techniques like image pre-processing, background subtraction, or better lighting conditions.
- **Background and Occlusion Handling:** Incorporating background subtraction or advanced segmentation techniques could help improve performance in cluttered environments or when body parts are occluded.

7. Scalability and Deployment:

- For wider accessibility, the app could be scaled and deployed on cloud platforms, allowing for processing of large datasets or enabling remote users to analyze their movements without requiring powerful local hardware.
- Developing a mobile app version could make the pose estimation tool more accessible, allowing users to perform pose tracking with just their smartphones.

8. Improved Visual Feedback:

- Enhance the user interface with more intuitive and informative visualizations, such as heatmaps of joint angles or real-time posture analysis. For example, the app could highlight areas of the body where posture needs improvement or where strain might occur.
- **Augmented Reality (AR) Integration:** Incorporating AR could allow users to visualize their pose in 3D space and get real-time feedback on their alignment, providing more immersive and engaging experiences.

By focusing on these improvements, the model could expand its scope and functionality, addressing both the technical and practical challenges associated with human pose estimation. This would make it more effective for a wide range of applications, from healthcare and sports to entertainment and interactive systems.

5.2 CONCLUSION:

The Human Pose Estimation project significantly contributes to the growing field of computer vision and artificial intelligence by enabling real-time, accurate detection of human body movements and postures. Using the MediaPipe Pose model, the project allows for easy integration into applications that require motion tracking, such as fitness apps, physical therapy monitoring, and virtual reality (VR) systems. By processing images and estimating human poses, the project can be leveraged to analyze human motion for various purposes like assessing physical health, improving sports performance, enhancing interactive gaming, and even assisting in security surveillance systems.

One of the key contributions is its accessibility via a simple, user-friendly Streamlit app interface, where users can upload images and receive instant pose estimation feedback. This system's ability to detect and visualize key body landmarks and connections (e.g., head, shoulders, elbows, knees, etc.) makes it a powerful tool for understanding human motion, whether for medical, athletic, or entertainment purposes. Additionally, the integration of MediaPipe ensures that the application remains efficient, with fast processing times and high-quality outputs.

The project's impact is particularly notable in healthcare and sports industries, where understanding body posture and movement is crucial for rehabilitation, injury prevention, and optimizing performance. For example, physiotherapists can use the tool to monitor patients' movements and tailor recovery exercises, while athletes can use it to refine their techniques. Furthermore, the project contributes to the broader AI and machine learning fields by exploring pose estimation models that can be extended to other domains like gesture recognition or human-computer interaction.

Overall, the Human Pose Estimation project showcases the intersection of AI, health, and entertainment, driving innovations in how we interact with technology, analyze human motion, and make informed decisions based on visual data.

REFERENCES

[1] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.

[2] *Alexander Toshev, Christian Szegedy*; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1653-1660.

[3] *Wanli Ouyang, Xiao Chu, Xiaogang Wang*; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2329-2336.

[4] Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm Abhay Gupta, Kuldeep Gupta, Kshama Gupta and Kapil Gupta National Institute of Technology, Kurukshetra Haryana, India.

[5] Human Pose Estimation Using Machine Learning | Project Showcase | Machine Learning Projects by code arts link(https://www.youtube.com/watch?v=wQRFU79iC0c)

[6] Human Pose Estimation in Space and Time Using 3D CNN Conference paper First Online: 24 November 2016.Agne Grinciunaite, Amogh Gudi, Emrah Tasli & Marten den Uyl Part of the book series: Lecture Notes in Computer Science ((LNIP,volume 9915))