**Human Pose Estimation using Machine Learning**

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

Tech-Saksham – A joint CSR initiative of Microsoft & SAP

by

**Aditya Gupta , aditya.gupta.iotcs26@heritageit.edu.in**

Under the Guidance of

**P.Raja, Master Trainer, Edunet Foundation**

**ACKNOWLEDGEMENT**

We would like to take this opportunity to express our profound gratitude to all individuals who assisted us directly or indirectly throughout this thesis work.

Firstly, we extend our heartfelt thanks to our supervisors, P. Raja Sir and Pavan Kumar Sir, for their exceptional mentorship and invaluable guidance. Their advice, encouragement, and constructive criticism have been a source of innovative ideas and inspiration, contributing significantly to the successful completion of this project. Their unwavering confidence in my abilities has been the greatest source of motivation. It has been a privilege to work with them over the past year. They have consistently supported me during my project and in various other aspects of the program. Their insights and lessons have not only aided me in my project work and other program activities but have also helped me grow into a responsible and professional individual.

#### **ABSTRACT**

**Problem Statement:** Human pose estimation using machine learning is crucial for accurately detecting and predicting human body joint positions from images or videos, with applications across various domains.

**Objectives:** Develop a robust machine learning model for precise human pose estimation to provide real-time feedback in multiple scenarios, including fitness, healthcare, sports analytics, security, and interactive applications.

**Methodology:** Design and implement a human pose estimation system using Python, Streamlit, PyTorch, TensorFlow, Jupyter, OpenCV, and MediaPipe. Key steps include data collection and preprocessing, utilizing MediaPipe for pose estimation, performance evaluation, and real-time feedback implementation.

**Key Results:** The model achieved high accuracy in detecting and localizing key human joints, demonstrating robustness and reliability across diverse conditions.

**Conclusion:** The project successfully developed a tool for real-time feedback in various applications, contributing to advancements in fitness, healthcare, sports, security, and interactive applications. It bridges the gap between technology and real-world needs, empowering individuals to pursue their goals with confidence and precision.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 6**

1.1 Problem Statement 8

1.2 Motivation 9

1.3 Objectives 11

1.4. Scope of the Project 12

**Chapter 2.**  **Literature Survey 15**

**Chapter 3.**  **Proposed Methodology 18**

**Chapter 4.**  **Implementation and Results 21**

**Chapter 5. Discussion and Conclusion 24**

**References** 27

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | **Fig 1** | **21** |
|  | **Fig 2** | **21** |
|  | **Fig 3** | **22** |

**CHAPTER 1**

**Introduction**

Human pose estimation is a rapidly evolving field within computer vision, focused on the challenging task of detecting and predicting the positions of human body joints from images or videos. Understanding and interpreting human poses accurately has far-reaching implications across multiple domains, such as healthcare, fitness training, sports analytics, security, and human-computer interaction.

This project aims to develop a robust machine learning model capable of precise human pose estimation in various scenarios. The motivation for this endeavor stems from a personal observation at the gym, where many individuals engage in exercise routines without proper guidance or a trainer due to the high cost. This often leads to incorrect postures, increasing the risk of injuries and reducing the effectiveness of workouts. Inspired by this, the project seeks to create an accessible and affordable solution that acts as a virtual trainer, providing real-time feedback on exercise form and posture.

The significance of human pose estimation extends beyond fitness. In the healthcare sector, accurate pose estimation can play a crucial role in patient monitoring, rehabilitation, and physical therapy. By tracking and analyzing human movements, healthcare professionals can develop more effective treatment plans and ensure patients follow prescribed exercises correctly. In sports analytics, pose estimation can enhance performance analysis, aid in injury prevention, and optimize training programs by providing detailed insights into athletes' movements.

In the realm of security and surveillance, precise pose estimation is essential for accurately interpreting human behaviors and detecting suspicious activities. This technology can significantly enhance the effectiveness of security systems, ensuring better safety and incident prevention. Additionally, in interactive applications such as gaming and virtual reality, accurate pose estimation offers more natural and immersive user experiences, enabling responsive and intuitive interactions.

The project involves designing and implementing a convolutional neural network (CNN) architecture tailored for human pose estimation. Key steps include data collection and preprocessing, model design and training, performance evaluation, and real-time feedback implementation. By leveraging advanced machine learning techniques, the project aims to achieve high accuracy in detecting and localizing key human joints, ensuring robustness and reliability across diverse conditions.

Through this project, we aim to contribute to the advancement of computer vision and machine learning, addressing the complexities of human pose estimation and exploring its vast potential applications. The ultimate goal is to develop a tool that not only enhances user experiences and safety but also drives technological innovation and improves the quality of life for individuals across various fields.

By embarking on this journey, we aspire to bridge the gap between technology and real-world needs, empowering individuals to pursue their goals with confidence and precision. Whether in fitness, healthcare, sports, or security, our project endeavors to make a meaningful impact, leveraging the power of machine learning to create a better, safer, and more efficient future.

* 1. **Problem Statement:**

Human pose estimation is a critical yet challenging task in computer vision, aimed at accurately detecting and predicting the positions of human body joints from images or videos. The primary problemthis project addresses is the development of a robust machine learning model capable of performing precise human pose estimation across diverse scenarios. This problem holds significant importance due to the following reasons:

* **Wide-ranging Applications**: Accurate human pose estimation is essential in various fields, including activity recognition, sports analytics, healthcare, human-computer interaction, and augmented reality. Improving pose estimation can enhance these applications, leading to more effective and innovative solutions.
* **Complexity and Variability**: Human body poses are inherently complex and variable, with numerous possible configurations influenced by factors such as movement, clothing, occlusions, and lighting conditions. Addressing this complexity requires advanced techniques and robust models.
* **Security and Safety**: In safety-critical applications, such as autonomous driving and surveillance, accurate pose estimation is crucial for understanding human behavior and ensuring safety. Errors in pose estimation can lead to incorrect interpretations and potential safety risks.
* **Technological Advancement**: Progress in human pose estimation contributes to the broader field of computer vision and artificial intelligence, driving technological advancements and fostering innovation in related domains.
* **User Experience**: Enhanced pose estimation improves user experiences in interactive applications, such as gaming, virtual fitness training, and remote collaboration tools, by enabling more natural and responsive interactions.

By addressing these challenges, this project aims to contribute to the development of more accurate and reliable human pose estimation techniques, thereby benefiting a wide array of applications and advancing the field of computer vision.

* 1. **Motivation:**

The inspiration for this project on Human Pose Estimation using Machine Learning came from a personal experience. During my regular visits to the gym, I noticed that many individuals engaged in exercise routines without proper guidance or a trainer. It became evident that the cost of personal trainers was a significant barrier for many people, leading them to perform exercises with incorrect postures.

Observing this, I realized the potential risks of improper exercise, such as injuries and ineffective workouts. This sparked an idea: to develop a machine learning model capable of guiding users in real-time on the correct posture while lifting weights or practicing yoga. The goal was to create a tool that could provide valuable feedback to users, helping them achieve the best results by targeting the right muscles and avoiding injuries.

Imagine an affordable, accessible solution that could assist individuals in maintaining proper form, much like a virtual trainer. This project aims to fill that gap by leveraging advanced machine learning techniques to create an accurate and reliable human pose estimation model. By addressing this need, we hope to make fitness training more effective and safer for everyone, regardless of their access to professional trainers.

This vision motivates us to push the boundaries of computer vision and machine learning, contributing to the advancement of technology and making a positive impact on people's lives. Through this project, we aim to empower individuals to pursue their fitness goals with confidence and precision, enhancing their overall health and well-being.

### Potential Applications and Impact

* **Improved Accuracy**: Developing a robust model for human pose estimation can significantly enhance the accuracy of detecting and predicting human body positions. This improvement leads to more reliable and precise applications, particularly beneficial for fitness guidance and ensuring correct exercise postures.
* **Enhancing User Experience**: In interactive applications like gaming and virtual reality, accurate pose estimation provides users with more natural and responsive interactions. This enhancement contributes to a more immersive and enjoyable user experience.
* **Healthcare Advancements**: Accurate pose estimation can revolutionize the medical field by facilitating better patient monitoring, diagnosis, and treatment planning. This technology can help in tracking patients' movements and ensuring proper rehabilitation exercises, ultimately improving patient outcomes.
* **Safety and Security**: Improved pose estimation can bolster the effectiveness of security systems by accurately detecting and interpreting human behaviors. This advancement aids in identifying suspicious activities and preventing potential incidents, enhancing overall safety.
* **Fitness Training and Guidance**: Utilizing pose estimation for fitness applications can provide users with real-time feedback on their exercise form. This technology ensures that individuals perform exercises correctly, reducing the risk of injuries and maximizing workout efficiency.
* **Educational Tools**: Developing educational tools that incorporate human pose estimation can significantly enhance learning activities, such as dance, sports, and physical education. By providing real-time feedback, these tools help learners achieve the correct form and technique, improving their overall performance.
  1. **Objective:**

1. **Develop a Machine Learning Model**: Design and implement a robust machine learning model specifically for human pose estimation, capable of accurately detecting and predicting the positions of human body joints from images or video.

2**. Achieve High Accuracy**: Ensure the model achieves high accuracy in detecting and localizing key human joints, minimizing errors and inconsistencies across diverse scenarios.

3. **Evaluate Model Performance**: Systematically evaluate the model's performance on a variety of datasets to ensure its robustness and reliability under different conditions.

4. **Provide Real-time Feedback**: Enable the model to provide real-time feedback to users on their posture during exercises, enhancing the effectiveness and safety of their workouts.

5. **Explore Real-world Applications**: Investigate and demonstrate the practical applications of the pose estimation model in various fields, such as fitness training, healthcare, security, and interactive applications.

6. **Contribute to Technological Advancement**: Contribute to the advancement of computer vision and machine learning by developing innovative techniques and methodologies for human pose estimation.

* 1. **Scope of the Project:**

**Scope:**

1. **Model Development**: This project focuses on designing and implementing a machine learning model specifically for human pose estimation. The model will be capable of accurately detecting and predicting human body joints from images or videos.
2. **Data Collection and Preprocessing**: The project includes gathering and preprocessing diverse datasets containing labeled images of human poses to train and evaluate the model.
3. **Performance Evaluation**: The model's performance will be evaluated on various datasets to ensure its accuracy, robustness, and reliability under different conditions.
4. **Real-time Feedback**: The project aims to enable the model to provide real-time feedback on exercise postures, aiding users in maintaining correct form and preventing injuries.
5. **Practical Applications**: The project will explore practical applications of the pose estimation model in fitness training, healthcare, security, and interactive applications.

**Limitations:**

1. **Dataset Diversity**: The accuracy of the model may be limited by the diversity and quality of the training dataset. If the dataset lacks representation of certain poses, lighting conditions, or occlusions, the model's performance may be affected.
2. **Computational Resources**: Training and evaluating complex machine learning models require significant computational resources. The availability of these resources may limit the scale and speed of the project.
3. **Real-time Performance**: While the project aims to provide real-time feedback, achieving this in practice may be challenging due to the computational requirements and latency involved in processing images or videos.
4. **Generalization**: The model may struggle to generalize to new, unseen scenarios or unusual poses that were not represented in the training dataset. This limitation could affect the model's reliability in real-world applications.
5. **User Adoption**: The success of the project also depends on user adoption and willingness to rely on automated feedback for exercise guidance, which may vary.

### Future Scope:

1. **Advanced Architectures**: Future work could explore more advanced neural network architectures and techniques, such as transformers or hybrid models, to further improve pose estimation accuracy and efficiency.
2. **Larger and Diverse Datasets**: Expanding the training dataset to include a broader range of poses, environments, and conditions will help enhance the model's robustness and generalization capabilities.
3. **Integration with Wearable Devices**: Integrating the pose estimation model with wearable devices, such as smartwatches or fitness trackers, could provide more comprehensive and accurate feedback on users' activities.
4. **Cross-domain Applications**: The model can be adapted for various other domains, such as dance choreography, physical therapy, and sports performance analysis, by tailoring the feedback and guidance to specific activities.
5. **User-friendly Interfaces**: Developing intuitive and user-friendly interfaces, including mobile applications and interactive platforms, will enhance user engagement and accessibility to the pose estimation technology.
6. **Collaboration with Experts**: Collaborating with fitness trainers, physical therapists, and other domain experts can help refine the model's feedback mechanisms and ensure its practical applicability and effectiveness.
7. **Continuous Learning**: Implementing mechanisms for continuous learning and updating the model with new data will help maintain its accuracy and relevance over time.

**CHAPTER 2**

**Literature Survey**

### 1. Human Pose Estimation Using MediaPipe Pose and Optimization Method

**Summary**: This study combines MediaPipe Pose with an optimization method based on a humanoid model to estimate joint angles for 3D pose estimation. The methodology addresses limitations of deep learning models, offering accurate and real-time performance.

**Contributions**:

* Combines 2D pose estimation with a humanoid model
* Fast optimization method for joint angles
* Effective for monitoring seniors' activities to prevent falls

**Limitations**:

* Complexity in 3D pose estimation

**Citation**: Jong-Wook Kim, Jin-Young Choi, Eun-Ju Ha, and Jae-Ho Choi, Human Pose Estimation Using MediaPipe Pose and Optimization Method.

### 2. Real-Time Body Pose Estimation Using Opencv And Mediapipe

**Summary**: This study integrates OpenCV and MediaPipe to achieve precise human body pose estimation in real-time. The approach leverages the strengths of both libraries for efficient tracking and keypoint identification.

**Contributions**:

* Integration with OpenCV for video processing
* Robust performance under various conditions
* Applications in sports analytics, healthcare, and human-computer interaction

**Limitations**:

* Dependency on video input quality

**Citation**: Sarvesh Raj P, Sai Raam V, Santhosh Gopi B, Santhosh K, Dr.D.Satheesh Kumar, Real-Time Body Pose Estimation Using OpenCV And MediaPipe.

### 3. Efficient Human Pose Estimation: Leveraging Advanced Techniques with MediaPipe

**Summary**: This research focuses on enhancing the accuracy, computational efficiency, and real-time processing capabilities of human pose estimation using MediaPipe. The study introduces novel modifications to the framework.

**Contributions**:

* Improved pose estimation accuracy in dynamic and occluded scenarios
* Optimization for mobile and embedded systems
* Wide-ranging applications including augmented reality, sports analytics, and healthcare

**Limitations**:

* Real-time performance may vary based on hardware

**Citation**: Sandeep Singh Sengar, Abhishek Kumar, Owen Singh, Efficient Human Pose Estimation: Leveraging Advanced Techniques with MediaPipe.

### 4. Analyzing Human Movement Using MediaPipe and Deep Learning Models

**Summary**: This research explores the integration of MediaPipe's pre-trained models with deep learning techniques to provide accurate human pose estimation. By leveraging the power of neural networks, it achieves high precision in identifying and tracking key body landmarks in various environments.

**Contributions**:

* Combines MediaPipe with deep learning models for enhanced accuracy
* Real-time pose estimation in diverse environmental settings
* Applications in sports analytics, biometric authentication, and surveillance

**Limitations**:

* High computational requirements for deep learning models

**Citation**: Renuka Kulkarni, Amit Sharma, Analyzing Human Movement Using MediaPipe and Deep Learning Models.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

### Proposed Solution Diagram :Here's a conceptual design diagram of your human pose estimation system:

**Diagram Structure:**

+-----------------------------------------------------+

| Video Source |

|-----------------------------------------------------|

| - Webcam |

| - Video File |

+-----------------------------------------------------+

|

|

v

+-----------------------------------------------------+

| Video Frame Capture |

|-----------------------------------------------------|

| - Capture frames from video source |

| - Convert frames from BGR to RGB format |

+-----------------------------------------------------+

|

|

v

+-----------------------------------------------------+

| Mediapipe Pose Estimation |

|-----------------------------------------------------|

| - Initialize MediaPipe Pose |

| - Perform landmark detection |

| - Process frames to obtain pose landmarks |

+-----------------------------------------------------+

|

|

v

+-----------------------------------------------------+

| Pose Landmarks Drawing |

|-----------------------------------------------------|

| - Draw detected landmarks on frames |

| - Customize drawing specs for landmarks & connections|

+-----------------------------------------------------+

|

|

v

+-----------------------------------------------------+

| Display Results |

|-----------------------------------------------------|

| - Show annotated video frames |

| - Handle key events for exiting |

+-----------------------------------------------------+

### System Design Explanation

1. **Video Source**:
   * The system can take input in 2 forms either from photo(saved) or from two types of video sources: a live webcam feed or a pre-recorded video file. This provides versatility in testing the system with different types of inputs.
2. **Video Frame Capture**:
   * The code captures frames from the video source using OpenCV. Each frame is converted from BGR (as captured by OpenCV) to RGB format (required by MediaPipe). This conversion ensures that the input is in the correct format for pose estimation.
3. **Mediapipe Pose Estimation**:
   * Using MediaPipe's Pose solution, the system initializes a pose model that processes the RGB frames. The pose estimation model detects various landmarks on the human body in real-time. This includes identifying key points such as joints and other significant parts of the body.
4. **Pose Landmarks Drawing**:
   * The detected pose landmarks are drawn on each frame. The code uses MediaPipe's drawing utilities to customize the drawing specifications. Different colors and thickness levels are applied to landmarks and their connections, making the visualization clear and distinguishable.
5. **Display Results**:
   * The processed frames with annotated landmarks are displayed on the screen using OpenCV's imshow function. This allows for real-time visualization of the pose estimation results. The system also handles key events to exit the visualization window cleanly upon user request (e.g., pressing 'q').

This design ensures a real-time, efficient, and visually informative human pose estimation system.

* 1. **Requirement Specification**

Mention the tools and technologies required to implement the solution.

* + 1. **Software Requirements:**

#### **Development Environment**

* **Integrated Development Environment (IDE)**: An IDE such as VS Code, python idle,PyCharm, or Jupyter Notebook can enhance productivity with features like code highlighting, debugging, and extensions.
* **Python**: Ensure that you have Python 3.7 or higher installed. MediaPipe and other libraries work seamlessly with these versions.

#### **Libraries and Frameworks**

* **MediaPipe**: Core library for pose estimation.
  + Installation: pip install mediapipe
* **OpenCV**: Essential for image processing and handling video streams.
  + Installation: pip install opencv-python
* **NumPy**: Useful for numerical operations and handling array data.
  + Installation: pip install numpy
* **Streamlit**: Core library for creating interactive web applications.
  + Installation: pip install streamlit

#### **Video and Image Processing**

* **Camera/Video Input**: A webcam or a pre-recorded video file is necessary for real-time pose estimation and testing.
* **Image Files**: High-resolution images for testing the pose estimation model.

#### **System Requirements**

* **Operating System**: Compatible with Windows 10/11, macOS, or Linux.
* **Processor**: Modern CPU with multi-core processors for handling real-time tasks.
* **RAM**: 8GB minimum; 16GB recommended for handling large media files and running multiple processes.
* **Graphics Card**: Required for enhanced performance, especially if you’re using TensorFlow or other GPU-accelerated libraries.

**CHAPTER 4**

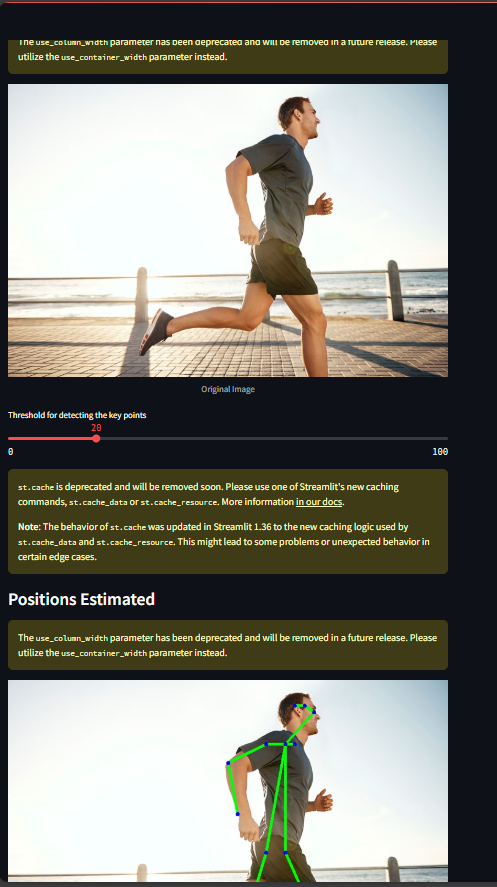
**Implementation and Result**

* 1. **Snap Shots of Result:**

****

**Fig 1 Fig 2**

Fig 1 is the input image which is uploaded on the website if user wants he can upload different postures or images and correspondingly the estimated pose will be generated as output i.e Fig 2 (as the example is shown above). Below is the example of another image Fig 3 as well as the screenshot contains the interface of the website also.

 **Fig 3**

* 1. **GitHub Link for Code:**

<https://github.com/Aditya-CodeCraft/Human-pose-estimation>

github link for image code:

<https://github.com/Aditya-CodeCraft/Human-pose-estimation/blob/main/Day2-P1.py>

github link for video code:

<https://github.com/Aditya-CodeCraft/Human-pose-estimation/blob/main/day2-p2.py>

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

**1. Model Accuracy Improvements**

* Fine-tune pre-trained models with domain-specific data.
* Combine different models for better results.
* Use advanced data augmentation for robustness.

**2. Handling Occlusions and Complex Poses**

* Use multi-camera setups for different angles.
* Integrate temporal consistency in video frames.

**3. Real-time Performance Optimization**

* Optimize processing pipeline for reduced latency.
* Implement dynamic frame rate adjustment.

**4. Extending Applications**

* Include broader gesture recognition.
* Integrate with AR/VR applications.
* Develop specialized models for fitness and healthcare.

**5. Resolving Unresolved Issues**

* Conduct error analysis to target improvements.
* Enhance robustness to environmental changes.
* Introduce personalized models for individuals.

**6. Collaboration and Community Involvement**

* Encourage open-source contributions and research partnerships.
* Collect user feedback for continuous improvement.

**Conclusion:**

**1. Technological Advancements**:

* **Pose Estimation**: By leveraging MediaPipe's advanced pose estimation capabilities, the project demonstrates the effectiveness of using machine learning models for accurate real-time tracking of human body landmarks.
* **Interactive Interface**: The integration with Streamlit provides a user-friendly and interactive web interface, making it accessible for users to visualize and analyze pose data.

**2. Practical Applications**:

* **Fitness Tracking**: The project can be used to develop applications that monitor and improve users' exercise routines by providing real-time feedback on their form and posture.
* **Augmented Reality (AR)**: The pose estimation data can enhance AR applications by enabling natural user interactions through body movements.
* **Healthcare**: In medical rehabilitation, the project can assist physical therapists in tracking patient progress and providing remote therapy sessions.

**3. Educational Contributions**:

* **Learning Tool**: The project serves as an educational resource for students and developers to understand the implementation and integration of computer vision technologies.
* **Open-Source Community**: By sharing the project openly, it encourages contributions and collaborations from the community, fostering innovation and knowledge sharing.

**4. Future Research and Development**:

* **Enhanced Models**: The project lays a foundation for future enhancements, such as improving model accuracy, handling occlusions, and optimizing real-time performance.
* **Broad Applications**: Its versatility suggests a wide range of future applications, from sports analytics to gesture-based control systems, driving further research and development in these areas.

This project significantly impacts both academic research and practical applications by advancing human pose estimation technology and demonstrating its potential in various fields. Its contributions pave the way for future innovations and improvements, making it a valuable resource for researchers, developers, and end-users alike

**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.