# 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

**Mahi Devendrabhai Patel, mahi.p@ahduni.edu.in**

Under the Guidance of

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

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who has supported me throughout this journey. Completing this thesis would not have been possible without the guidance, encouragement, and kindness of many individuals.

I am especially grateful to **P. Raja**, Master Trainer at Edunet Foundation, for his invaluable mentorship and constant support. His insightful advice, patience, and unwavering belief in my abilities have been instrumental in shaping this work. His guidance extended beyond just academics, inspiring me to think critically, explore new ideas, and develop a strong sense of professionalism.

I would also like to extend my appreciation to my peers, friends, and family, who have been my pillars of strength. Their encouragement and understanding helped me stay motivated and focused throughout this journey.

This thesis is a result of collective efforts, and I am truly thankful to each and every person who played a part in making it a reality.

## ABSTRACT

Provide a brief summary of the project, including the problem statement, objectives, methodology, key results, and conclusion. The abstract should not exceed 300 words.

Human pose estimation is a vital task in computer vision with applications in activity recognition, motion tracking, and human-computer interaction. This project focuses on implementing a real-time human pose estimation system using OpenCV and pre-trained deep learning models.

The problem addressed in this work is accurately detecting and visualizing human body key points in images and videos. Instead of training a model from scratch, we utilize a pre-trained model to identify key joints such as the head, shoulders, elbows, knees, and ankles. These points are then connected to form a skeletal representation of the human body.

The main objectives of this project are:

1. Implementing a human pose estimation system using OpenCV and a pre-trained model.

2. Detecting and marking key body joints on images and video streams.

3. Connecting the detected points to create a skeletal structure for pose estimation.

The methodology involves processing input frames through a pre-trained deep learning model that predicts body joint locations. OpenCV is then used to overlay these key points and draw lines between them to form a structured human pose. The system is tested on various images and videos to evaluate its performance in real-world scenarios.

Key results show that the pre-trained model efficiently detects human poses in different environments with good accuracy and real-time performance. The approach is effective for applications such as activity monitoring and gesture recognition. However, challenges such as occlusions and variations in body positions can affect accuracy.

In conclusion, this project successfully integrates a pre-trained model with OpenCV to estimate and visualize human poses. The results demonstrate the feasibility of using deep learning for real-time pose detection, opening doors for further applications in sports analysis, virtual reality, and security surveillance.

# TABLE OF CONTENT

# LIST OF FIGURES

# CHAPTER 1

# Introduction

Human pose estimation is a computer vision technique used to detect and analyze body posture by identifying key joints in images and videos. In fitness training, maintaining the correct posture is crucial to prevent injuries and ensure effective workouts. However, trainers cannot always monitor every trainee in real-time, making it challenging to provide immediate feedback.

This project utilizes a pre-trained deep learning model with OpenCV to estimate human poses accurately. The system detects key body points, connects them to form a skeletal representation, and visualizes movement in real-time. By automating posture analysis, this project aims to assist trainers and trainees in improving exercise form, reducing injury risks, and enhancing training efficiency in both gym and home workout environments.

## 1.1 Problem Statement

In a gym or fitness environment, trainers often struggle to monitor every trainee simultaneously, making it difficult to provide real-time feedback on posture and exercise form. Incorrect posture during workouts can lead to ineffective training and an increased risk of injury. Traditional manual supervision is limited and may not be feasible for large groups or virtual training sessions. This project addresses this problem by implementing a human pose estimation system using a pre-trained deep learning model and OpenCV. The system detects key body joints, connects them to form a skeletal representation, and helps analyze posture in images and videos, making it a valuable tool for fitness tracking and automated workout assessments.

## 1.2 Motivation

While working out, I often struggled with maintaining the correct posture, especially without a trainer’s constant supervision. In a busy gym, trainers can't monitor every trainee in real-time, increasing the risk of improper form and potential injuries. This challenge inspired me to develop an automated system that uses machine learning and OpenCV to analyze human posture, helping both trainers and trainees improve exercise form, prevent injuries, and enhance workout efficiency—whether in a gym or at home.

## 1.3 Objectives

The main objectives of this project are:

- To develop a human pose estimation system using OpenCV and a pre-trained deep learning model.

- To detect and visualize key body joints in images and video streams.

- To connect detected points to create a skeletal structure for pose analysis.

- To explore the system's potential for fitness monitoring and posture correction.

## 1.4 Scope of the Project

This project focuses on 2D human pose estimation using a pre-trained deep learning model. The system is designed to work in real-time on images and videos, helping trainers and trainees analyze body posture. However, limitations include challenges in detecting poses under poor lighting conditions, occlusions, and complex body movements. Future work can extend this to 3D pose estimation and integration with AI-driven feedback mechanisms for automated posture correction. Project can be used in Yoga centers, sport field gym anywhere.

# CHAPTER 2

# Literature Survey

## 2.1Review relevant literature or previous work in this domain.

**Review of Relevant Literature on Human Pose Estimation Using Deep Learning**

Human Pose Estimation (HPE) is a critical task in computer vision aimed at predicting the location of human joints from images and videos. It has applications in areas like sports analysis, surveillance, and healthcare. Deep learning has significantly advanced HPE by improving accuracy and robustness against challenges like occlusions and crowded scenes. [1] The systematic review by Samkari et al. (2023) comprehensively analyzed HPE research, covering methods for both images and videos and single-person and multi-person detection

## 2.2Mention any existing models, techniques, or methodologies related to the problem.

The literature identifies two main approaches: traditional computer vision-based methods and deep learning-based models. Traditional methods relied on handcrafted features, such as Histogram of Oriented Gradients (HOG) and pictorial structures, but struggled with complex poses and occlusions. Deep learning-based methods, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown superior performance. Key advancements include:

- DeepPose: The first deep learning-based HPE model that used CNNs to regress joint positions directly.
- OpenPose, DeepCut, and Mask R-CNN: Multi-person pose estimation frameworks using top-down and bottom-up approaches.
- Pretrained models and loss functions: Techniques such as multi-stage refinement, heatmap-based regression, and loss function optimizations have enhanced pose estimation accuracy.

## 2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

Despite advancements, challenges remain:

- Occlusions and crowded scenes: Many models struggle to detect overlapping joints in dense environments.
- Computational complexity: High-accuracy models require substantial computational power, limiting real-time applications.
- Limited datasets: While datasets like COCO and MPII provide extensive annotations, there is a need for more diverse and large-scale datasets with better occlusion handling.
- Generalization issues: Many models perform well on benchmark datasets but struggle with unseen environments.

Addressing the Gaps

Future research should focus on:

- Improving robustness: Developing occlusion-resistant models through self-supervised learning and transformer-based architectures.
- Reducing computational load: Implementing lightweight models that balance accuracy and efficiency.
- Enhancing datasets: Incorporating more diverse human activities and real-world scenarios for better generalization.

1. Samkari, E., Arif, M., Alghamdi, M., & Al Ghamdi, M. A. (2023). Human pose estimation using deep learning: a systematic literature review. *Machine Learning and Knowledge Extraction*, *5*(4), 1612-1659.

# CHAPTER 3

# Proposed Methodology

## 3.1 System Design

Provide the diagram of your Proposed Solution and explain the diagram in detail.

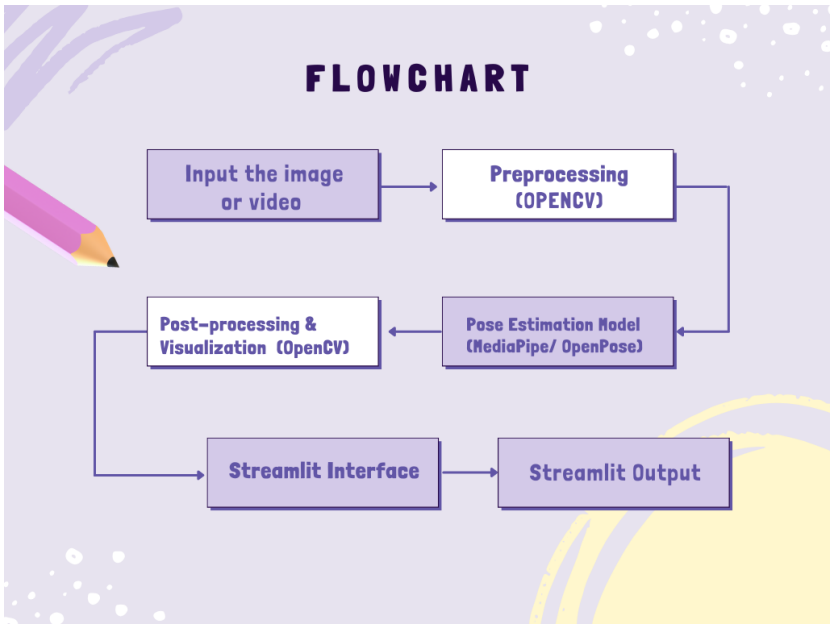


*Figure 1Flowchart*

Input (Image/Video Stream): The system receives input from an image or live video feed through a camera.

Preprocessing: OpenCV processes the input to adjust resolution and format, making it compatible with the pose estimation model.

Pose Estimation Model: A pre-trained model (such as MediaPipe or OpenPose) detects key body joints and outputs their coordinates.

Post-processing & Visualization: OpenCV is used to connect the detected points and create a skeletal structure over the body in the image/video.

Streamlit Interface: Streamlit is used to create an interactive frontend that allows users to upload images/videos, display the pose estimation results, and view real-time feedback on posture.

Output Display: The processed images/videos with overlaid pose estimations are displayed via the Streamlit interface for analysis.

## 3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

### 3.2.1 Software Requirements:

- Operating System: Windows, Linux, or macOS
- Programming Language: Python
- Libraries & Frameworks:
    - OpenCV (for image processing and visualization)
    - NumPy (for handling numerical operations)
    - MediaPipe / OpenPose (pre-trained models for pose estimation)
    - Streamlit (for creating the interactive web interface)
- IDE/Text Editor: VS Code, Jupyter Notebook, or PyCharm
- Dependencies: pip (for installing required libraries)

# CHAPTER 4

# Implementation and Result

## 4.1 System Implementation

The implementation of the Human Pose Estimation using Machine Learning project involves integrating a pre-trained deep learning model with OpenCV to detect and visualize human body key points in images and videos. The following steps outline the system implementation:

Data Input

- The system accepts input in the form of images or live video streams from a webcam or uploaded files.
- OpenCV is used to process the input for compatibility with the pose estimation model.

Preprocessing

- Images and video frames are resized and converted to the required format before being fed into the model.
- OpenCV handles noise reduction, contrast adjustment, and other preprocessing techniques.

Pose Estimation Model

- A pre-trained deep learning model (such as MediaPipe, OpenPose, or DeepPose) is used to identify key body joints.
- The model processes the input and outputs the coordinates of the detected body joints.

Post-processing & Visualization

- The detected key points are connected to create a skeletal structure of the human pose.
- OpenCV is used to overlay these points and lines on the original image or video.

Interactive Interface with Streamlit

- The system is deployed using Streamlit, providing a user-friendly interface for uploading images and viewing results.

- Users can adjust parameters like threshold values to fine-tune detection accuracy.

## 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.

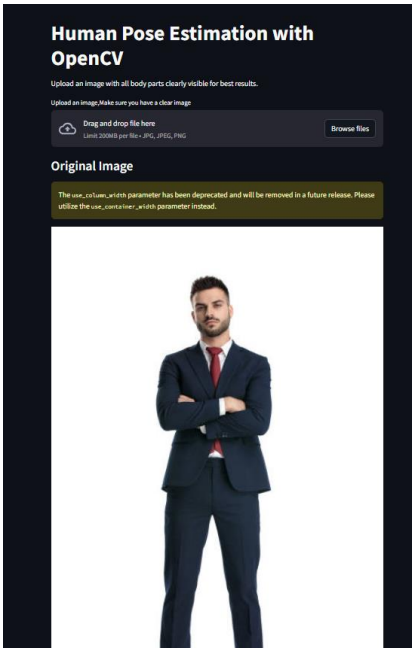


*Figure 2Example 1*


*Figure 3 Example 2*

First, browse and upload the image or video. Then, set the threshold within a range of 0 to 100.

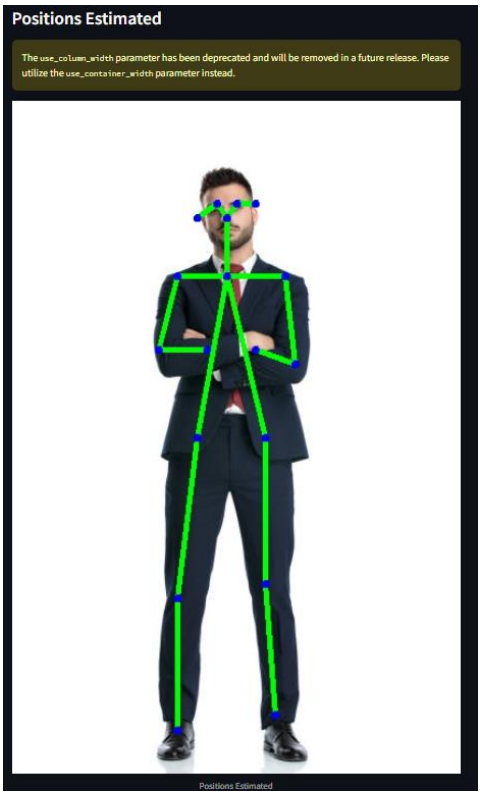


*Figure 4 Pose Estimation*

The image displays the estimated position, highlighting key landmarks and the skeletal structure.

### 4.2 GitHub Link for Code:

https://github.com/Mahi3112/AICTE-Internship.git

# CHAPTER 5

# Discussion and Conclusion

## 5.1 Future Work:

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

Future improvements could focus on:

Enhancing the model's accuracy in challenging scenarios, such as occlusions and poor lighting.

Expanding to 3D pose estimation for more precise movement analysis.

Adding real-time feedback for posture correction during exercises.

Supporting multi-person detection to cater to gym environments with multiple trainees.

Integrating with wearable devices to provide a comprehensive fitness tracking system.

## 5.2 Conclusion:

Summarize the overall impact and contribution of the project.

This project developed a human pose estimation system using OpenCV, a pre-trained model, and Streamlit for an interactive interface. The system helps improve workout form and reduce injury risks by analyzing body posture in real-time. It offers significant contributions to AI-driven fitness applications and can be further expanded to support more complex use cases like multi-person tracking and 3D pose estimation, making it useful in various fields such as sports and rehabilitation.

# REFERENCES

1. Samkari, E., Arif, M., Alghamdi, M., & Al Ghamdi, M. A. (2023). Human pose estimation using deep learning: a systematic literature review. *Machine Learning and Knowledge Extraction*, *5*(4), 1612-1659.