Human Pose Estimation

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

Sonali Narayan Patil

Sonalipatil0725@gmail.com

Under the Guidance of

Mr. P. Raja

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who contributed to the successful completion of this project.

First and foremost, I extend my heartfelt appreciation to **Honourable Mr. P. Raja Sir** my project mentor, for their invaluable guidance, constant support, and constructive feedback throughout the development of this project. Their expertise and encouragement have been instrumental in shaping this work.

I am also thankful to **TechSaksham Team** for providing the necessary resources and a conducive environment for carrying out this project.

Lastly, I express my deep gratitude to team for their motivation, patience, and encouragement throughout this journey. Their support has been my driving force in completing this project successfully.

Thank you all for your contributions and support.

Ms. Sonali Narayan Patil

ABSTRACT

The goal of the Human Pose Estimation Project is to create a system that can recognize and forecast important human body parts from pictures or video, such as joints like the elbows, knees, and shoulders. The project focuses on developing an effective posture estimation system appropriate for a range of applications, such as fitness tracking, gesture recognition, and human-computer interaction, by utilizing sophisticated machine learning algorithms and pre-trained models.

Accurate body joint tracking and identification, real-time processing of video streams or live camera feeds, and user-friendly visual overlays for position skeletons and identified important points are some of the project's salient characteristics. These features allow the system to be used in animation, sports analysis, activity identification, and rehabilitation.

To provide accurate and dependable posture estimation, the methodology uses cutting-edge machine learning models that have been trained on large datasets. The system achieves real-time performance by combining these models with effective processing pipelines, which makes it suitable for interactive applications.

In conclusion, this project shows that it is possible to develop a reliable human pose estimate system that combines precision with real-time functionality. Its adaptability to a variety of sectors underscores its potential to improve human-computer interactions and offer insightful information in industries like healthcare and sports.

TABLE OF CONTENT

Abstract I

Chapter 1. Introduction 1

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 2

Chapter 2. Literature Survey 3

Chapter 3. Proposed Methodology 7

Chapter 4. Implementation and Results 10

Chapter 5. Discussion and Conclusion 13

References 15

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Fig 1	**Top-down approach and bottom-down approach**	**5**
Fig 2	**Human Pose Estimation**	**7**
Fig 3	**Pose Landmarks**	**12**
Fig 4	**Pose Drawing**	**13**
Fig 5	**Pose Estimation of Video**	**14**

CHAPTER 1

Introduction

1.1 Problem Statement:

Accurate human pose estimation is challenging due to variations in poses, occlusions, and diverse imaging conditions. Traditional methods struggle with real-time processing and generalization. This project develops an efficient system using advanced machine learning to achieve accurate, real-time detection and visualization of human body key points.

Significance of the Problem:
Estimating human pose is essential for several applications, such as:

1. Activity Recognition: Recognizing human motions and behaviours for behavioural study, human-computer interaction, and monitoring.
2. Sports analysis: By examining body mechanics, coaches and athletes can assess and enhance performance.
3. Animation and gaming: Using human gestures to create lifelike character animations.
4. Rehabilitation: Tracking progress and ensuring proper exercise execution by keeping an eye on patients' movements during physical treatment.

1.2 Motivation:

Human pose estimation is becoming more and more popular because of its many uses, which is what inspired this research. The project aims to contribute to domains like fitness tracking, where real-time feedback on exercise form can improve individual training experiences, by creating an effective and precise pose estimate system. Furthermore, improvements in gesture recognition can enhance human-computer interactions, increasing the usability and accessibility of technologies. The project's emphasis on real-time processing and visualization is in line with the growing need across a range of industries for applications that are responsive and easy to use.

Potential Applications and Impact:
Human pose estimation has a variety of real-life applications, including:

1. **AI-powered personal trainers:** Apps can improve at-home training regimens by offering comprehensive real-time feedback on exercise form.

2. **Professional sports:** While judges gain from objective evaluations, athletes can enhance their performance by carefully examining their approaches.
3. **Filmmaking and gaming:** By better capturing and representing human movements, character animation and interactive gaming experiences can be improved.
4. **Healthcare:** By offering precise evaluations of body posture and motion, pose estimation can support physical therapy and rehabilitation, resulting in individualized treatment regimens and better patient outcomes.

1.3 Objective:

1. **Key Point Detection:** Recognize and monitor key human body joints precisely to comprehend human postures.
2. **Real-Time Processing:** To enable real-time pose estimation, efficiently analyse live camera feeds or video streams.
3. **Visualization:** Offer user-friendly visual overlays that show pose skeletons and identified important areas on pictures or videos.
4. **Application Development:** By offering trustworthy posture estimation data, you may make it easier to create applications in fields like activity identification, sports analysis, animation, and rehabilitation.

1.4 Scope of the Project:

1.4.1 Scope:

1. Find and follow the main features of the human body in pictures or videos.
2. Make it possible to process live camera feeds in real time. applications in sports analysis, rehabilitation, and fitness.

1.4.2 Limitations:

1. Processing videos in real time could necessitate a large amount of processing power.
2. Environmental variables and image quality can affect accuracy.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature

In computer vision, Human Pose Estimation (HPE) is a crucial problem that entails estimating the spatial locations of important body joints from visual inputs like pictures or movies. Deep learning and the availability of large-scale datasets have combined to generate notable breakthroughs in this field during the past 20 years. Human-computer interaction, gaming, sports analytics, fitness tracking, virtual reality, and healthcare (including rehabilitation and patient monitoring) are just a few of the many and varied fields in which HPE finds use.

Early Research and Conventional Methods: Manual feature extraction and geometric rule modeling of human body structures were the mainstays of the early HPE techniques. In order to model spatial relationships, early techniques such as the Pictorial Structure Model (PSM) depicted the body as a group of components joined by springs. Although fundamental, these techniques had issues with accuracy and scalability in complicated, multi-person situations.

2.1.1 Deep Learning Revolution: With techniques like these, the advent of deep learning revolutionized HPE.

1. One of the first to predict joint coordinates using convolutional neural networks (CNNs) was Toshev and Szegedy's DeepPose (2014). Pose estimation was handled as a regression problem by this model, which produced notable accuracy gains.
2. Stacked Hourglass Networks (2016): Presented a new architecture that improved position predictions at various resolutions by switching between downsampling and upsampling. When it comes to single-person pose estimation, this paradigm became the industry standard.
3. A bottom-up method called OpenPose (2017) found important points for each person in a picture and categorized them into postures. In situations involving multiple people, it worked especially well.

2.1.2 Datasets Driving Progress: Important datasets have propelled HPE's development.

1. With more than 200,000 annotated keypoints, COCO (Microsoft Common Objects in Context) is a standard for multi-person HPE.
2. The MPII Human Pose Dataset offers a platform for both single-person and multi-person pose estimate assessments by showcasing difficult positions taken in real-world situations.
3. PoseTrack and AI Challenger enhanced HPE for dynamic and crowded contexts by introducing films and crowd-specific scenarios.

2.2 Existing Models, Approaches, and Methodologies

HPE techniques can be broadly divided into two approaches: top-down and bottom-up. Each has pros and cons of its own.

2.2.1 Top-Down Methods:

These algorithms estimate each person's pose within the bounding box after first identifying individuals in an image using object identification techniques like YOLO or Faster R-CNN. Notable instances consist of:

1. **SimpleBaseline (2018):** Uses deconvolution layers to forecast heatmaps for keypoints after extracting features from bounding boxes using a ResNet backbone.
2. **HRNet (2019):** Achieves state-of-the-art posture estimation results by maintaining high-resolution representations across the network.

2.2.2 Bottom-Up Methods:

These techniques identify every important feature in a picture and classify them into distinct poses. Important instances consist of:

1. **OpenPose:** Recognizes bodily parts and links them to specific people using Part Affinity Fields (PAFs).
2. **DeepCut and DeeperCut:** Concentrate on grouping identified keypoints to optimize multi-person pose estimation.

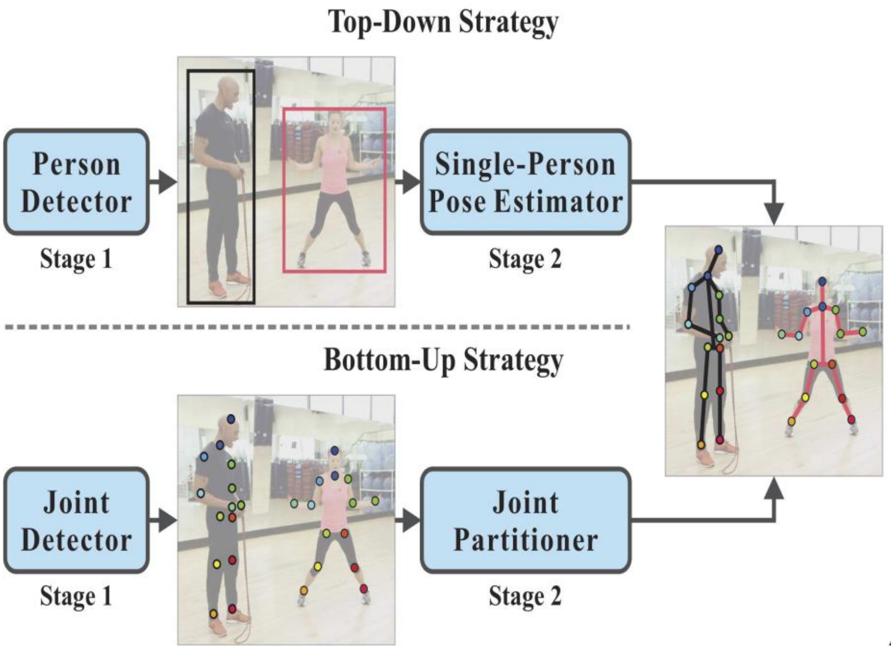

Fig1. Top-down approach and bottom-down approach

Advanced Techniques and Techniques:

1. **Attention Mechanisms:** To improve keypoint localization, techniques such as HRNet incorporate attention modules to concentrate on pertinent areas of the image.
2. **Transformers in HPE:** To describe long-range interdependence between joints, recent works, like PoseFormer, use transformers.

Keypoint Detection and Association:

The process of detecting keypoints and associating them with individuals involves techniques like confidence maps, affinity fields, and graph-based approaches.

Benchmarks & Datasets:

MSCOCO and MPII are two examples of datasets that are now used as standards for assessing HPE models. These datasets contain difficult situations like extreme stances, crowd scenes, and occlusion.

2.3 Gaps, Limitations, and Addressing Solutions:

2.3.1 Limitations:

Despite impressive advancements, HPE still faces a number of obstacles:

1. Occlusion and Complex Pose: When body parts are obscured by occlusions, many HPE models have trouble. For example, keypoint localization is not reliable when body parts are obscured by objects or in busy surroundings. Existing models are additionally challenged by complex positions, as those in yoga or acrobatics.

2. Real-Time Performance: On devices with limited resources, such smartphones, high-performing models are inappropriate for real-time applications since they frequently demand significant processing resources.

3. Lack of Contextual Understanding: Keypoints are frequently treated separately by models, which overlook the innate connections between joints. This can result in stances that are physiologically unrealistic, particularly in difficult situations.

2.3.2 Proposed Solutions in the GitHub Repository: The GitHub repository addresses these gaps by:

1. **Improving Occlusion Robustness:** Effectively models joint interdependence by utilizing techniques like compositional token-based representations and attention mechanisms.
2. **Optimizing Computational Efficiency:** To allow real-time performance, lightweight architectures and effective inference pipelines are used.
3. **Leveraging Augmented Data:** To enhance generalization across various stances and situations, this method makes use of synthetic augmentation techniques and a variety of datasets.

CHAPTER 3

Proposed Methodology

3.1 System Design

The goal of the Human Pose Estimation Project is to identify and forecast important human body parts from pictures or movies, including joints like the elbows, knees, and shoulders. Applications for this technology include human-computer interface, gesture recognition, and fitness tracking.

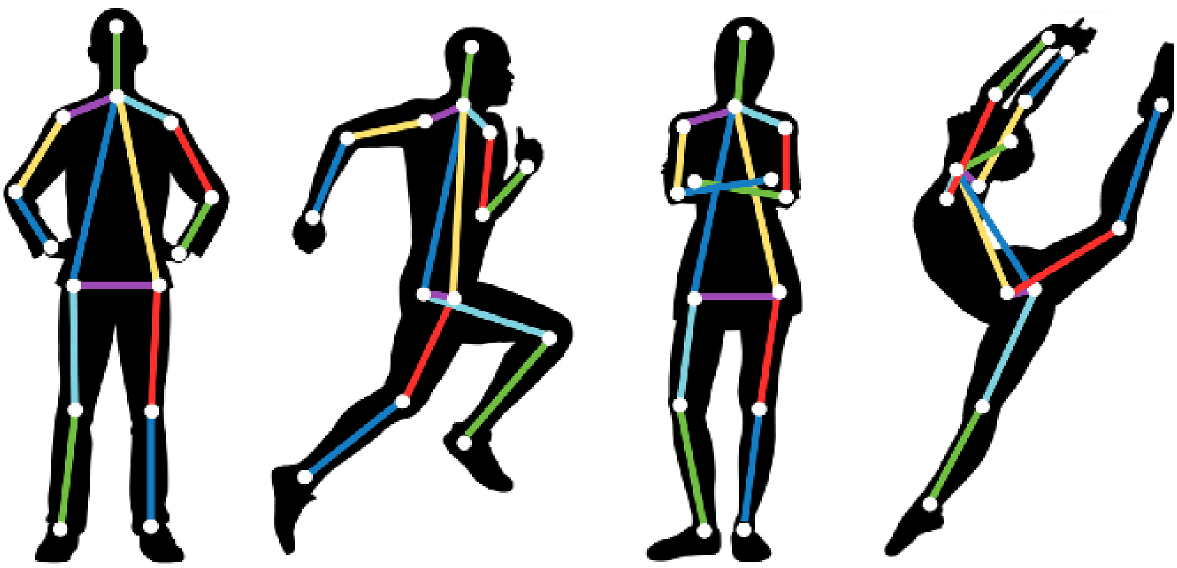

Fig2. Human Pose Estimation

3.2.1 Proposed Methodology:

1. **Data Collection and Preprocessing:**
    - **Dataset Acquisition:** Compile a large collection of pictures or videos that show people in various poses with annotations. One can use publicly accessible datasets like as Human3.6M, MPII, or COCO.
    - **Data Augmentation:** To increase the diversity and resilience of the dataset, use methods like rotation, scaling, flipping, and color modifications.

2. **Model Selection and Training:**
   - **Use of Pre-trained Models:** Use a pre-trained model as the basis, such as PoseNet or OpenPose. These models have proven to be effective at challenges involving the estimate of human stance.
   - **Model Fine-Tuning:** Adjust the parameters of the previously trained model to fit the particular dataset, guaranteeing increased performance and accuracy.

3. **Pose Estimation Process:**
   - **Feature Extraction:** Use the model to process incoming video frames or images and extract pertinent information.
   - **Key Point Detection:** Using the features that were retrieved, determine and forecast the locations of important body joints.
   - **Pose Construction:** Create a skeletal depiction of the human stance by joining the identified essential points.

4. **Post-Processing and Visualization:**
   - **Confidence Scoring:** To evaluate the accuracy of the forecasts, give each identified key point a confidence score.
   - **Visualization:** For a clear visual representation, place the skeletal model over the original pictures or video frames.

5. **Application Integration:**
   - **Real-Time Processing:** To effectively manage live video streams or camera feeds, optimize the system for real-time processing.
   - **Application Deployment:** Include the posture estimation system in programs such as sports analysis, animation, activity recognition, or rehabilitation aids.

3.2.2 Proposed Architecture:

The Human Pose Estimation system's architecture is organized as follows:

1. **Module of Input:**

   Accepts input in the form of pictures or video frames.
2. **Module for Preprocessing:**

   Prepares the model's input data by performing data standardization and augmentation.
3. **Feature Extraction Module:**

   Takes the input data and uses a convolutional neural network (CNN) to extract useful features.
4. **Module for detecting key points:**

   Uses a heatmap or regression-based method to forecast the locations of important body joints.
5. **Module for post-processing:**

   Refines the identified critical points using methods such as non-maximum suppression.
6. **Module for Visualization:**

   Overlays the identified poses onto the initial input data to create visual representations.
7. **Module for the Application Interface:**

   Offers a way to incorporate the posture estimate results into other applications.

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

1. **Opertaing System:** Latest Windows version or Linux- Ubuntu 16.04
2. **RAM -** 4 or more than 4 GB preferable

3.2.2 Software Requirements:

1. IDLE (Python 3.8 64-bit)
2. OpenCV library for python
3. Media pipe framework
4. Streamlit Open Source Python Framework

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Fig3. Pose Landmarks

1. This image also detects key pose landmarks, but instead of a skeleton representation, it uses green dots to highlight specific body points.
2. The model focuses on detecting individual keypoints without connecting them with lines.
3. This method is useful for feature extraction in tasks like gesture recognition and activity classification

Fig4. Pose Drawing

1. This image shows a human figure with red dots at key joint locations (e.g., head, shoulders, elbows, knees, etc.).
2. The joints are connected with white lines, forming a skeleton structure that represents the person's posture.
3. This is a typical pose estimation model output where a pre-trained deep learning model like OpenPose, Mediapipe, or HRNet detects key joints.
4. The person's arms are crossed, and the skeleton accurately captures this by showing elbows bent inward.

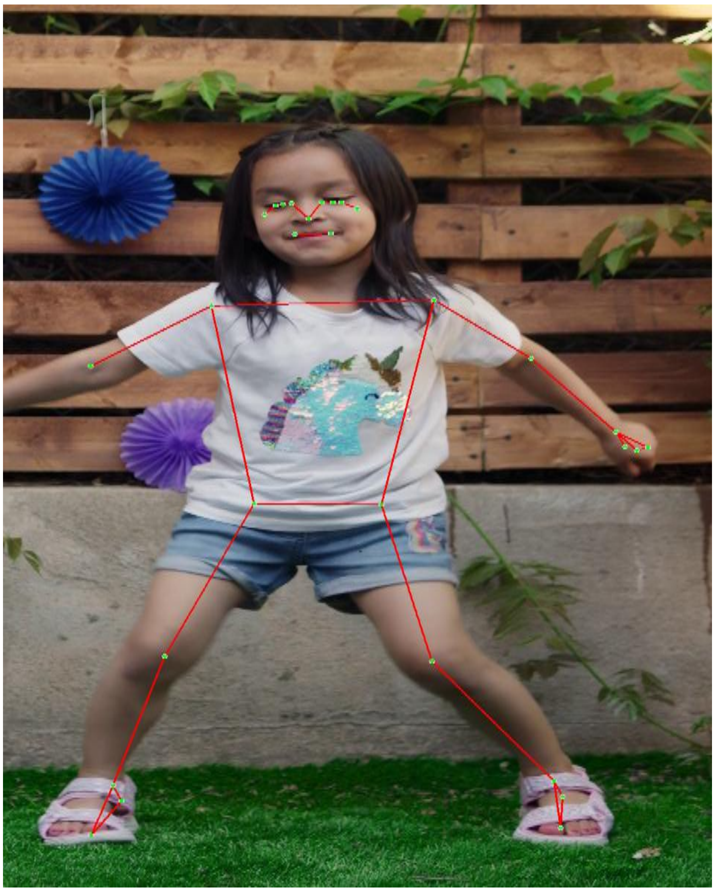

Fig5. Pose Estimation of Video

1. This snapshot is the output of video that applies pose estimation to a child in a dynamic pose.
2. It uses red lines to create a skeleton representation similar to the first image.
3. The detected pose shows the child standing with arms and legs spread out, which is accurately captured by the keypoints and skeleton lines.
4. The background consists of a natural setting, indicating that the model can work in real-world environments.

GitHub Link for Code:

https://github.com/Namu0725/Human-Pose-Estimation-Project

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

The following future scope points can be added to the "Human Pose Estimation" project to further expand it:

1. **Increased Accuracy:** By using more sophisticated models or optimizing current ones, including by adding more varied datasets, you can increase the accuracy of keypoint detection.
2. **Multi-Person Pose Estimation:** For applications in congested settings, extend the project to manage several people in a single frame.
3. **Real-Time Gesture Recognition:** Incorporate real-time gesture recognition into tasks involving interaction, including gadget control.
4. **Deep Learning Optimization:** Use optimization strategies to increase processing speed and efficiency while lowering latency in real-time.
5. **Cross-Platform Integration:** To make the system more accessible, create web-based or mobile versions.
6. **3D posture Estimation:** Use 3D posture estimation to enhance virtual reality and sports analysis by more accurately simulating the motions of the human body in space.

5.2 Conclusion:

The Human-Pose-Estimation-Project efficiently tracks important human body parts in both pictures and videos, including the head, shoulders, elbows, wrists, hips, knees, and ankles. The project guarantees great accuracy and resilience in a variety of settings by utilizing pre-trained models and sophisticated computer vision techniques.

Among its many notable accomplishments is its real-time data processing capability, which makes it perfect for applications like fitness tracking, gesture recognition, augmented reality, and interactive gaming that demand immediate response. By incorporating user-friendly visualization approaches, the project improves interpretability and usability by enabling users to view the identified important features and skeletal structures superimposed on the input photos or video frames. For researchers and developers working in fields including animation, sports analytics, medical rehabilitation, and human-computer interface, this makes it an invaluable tool.

Even with its success, there is still opportunity for development, such as strengthening the model's resistance to occlusions and changing lighting and making it more suitable for deployment on edge devices. It has a lot of promise for usage in surveillance, automation, and medical applications with more development.

REFERENCES

[1]. Prasad Dhore, Aparna Pande, Shital Mehta, Saili Sable, "Human Pose Estimation and Classification: A Review", NeuroQuantology, Volume 20, Issue 15, November 2022, Pages 3199-3213, DOI: 10.14704/NQ.2022.20.15.NQ88313.

[2]. Srividya Inampudi, Romik Amipara, Rutvik Kokate, "Human Pose Detection and Estimation", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 9, September 2020, ISSN: 2320-2882.

[3]. Liangchen Song, Gang Yu, Junsong Yuan, Zicheng Liu, "Human Pose Estimation and Its Application to Action Recognition: A Survey", Journal of Visual Communication and Image Representation, Elsevier, 2021.

[4]. Zigang Geng, Chunyu Wang, Yixuan Wei, Ze Liu, Houqiang Li, Han Hu, "Human Pose as Compositional Tokens", arXiv preprint, 2023.