Human Pose Detection: A Machine Learning Based Approach with Keypoint Extraction and Ensemble Methods

Harshitha Polsani, Venkata Ramya Vardineni, Anusha Yarrasani,

Dr. ABM Rezbaul Islam
Sam Houston State University, USA

Department of Computer Science
COLLEGE OF SCIENCE & ENGINEERING TECHNOLOGY
SAM HOUSTON STATE UNIVERSITY
MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

Introduction

• Human pose detection is primarily used in the research area of computer vision.

• Human pose detection is an important task in computer vision, with applications in fields such as sports and fitness tracking, health monitoring, and human-computer interaction.

• In many applications, human pose detection is used as a first step to extract useful information about the human body from images or videos.

• Our work specifically focuses on single person pose detection, which involves detecting the key points of a single person's body posture from an image or video.

• Our work aims to develop a machine learning-based approach for human pose detection using Mediapipe to extract keypoints and supervised learning on a comprehensive dataset..

• We merged two different datasets: a primary dataset of yoga poses images and a Yoga-82 dataset, which includes six poses that are not present in the primary dataset. These additional poses were obtained through web scraping.

Contributions

• We collected two different datasets, including a primary dataset of yoga pose images and a Yoga-82 dataset, to create a comprehensive dataset for human pose detection.

• We utilized Mediapipe to extract 33 key points from the images and addressed class imbalance issues by undersampling and oversampling of the landmark data.

• We trained various machine learning models such as Support Vector Machine(SVM), Random Forest(RF), K Nearest Neighbors(KNN), Bagging Classifier with different base estimators and ensembles of different combinations of these classifiers, as well as deep learning-based approaches like CNN and Feed Forward Neural Networks for pose detection.

• In addition to extracting landmark data, we explored a deep learning-based approach by using AlexNet directly on images, and performed data augmentation to address class imbalance issues..

• We implemented real-time pose detection and developed a web interface for users to upload images for pose detection.

Proposed Method

- **An Approach for Human Pose Detection using Machine Learning and Keypoint Extraction Using Mediapipe**
- **Data Collection**
  - We created a comprehensive dataset by merging two different datasets, resulting in approximately 6219 instances after invalid file formats were removed and the images were processed using Mediapipe.
- **Keypoint Extraction**
  - The pre-trained Mediapipe model was used to extract 33 key points from the images. The landmark data obtained using mediapipe pose detection model is having 6219 instances and 132 features which includes x,y,z and visibility coordinates of 33 keypoints. We performed undersampling of majority classes and oversampling of minority classes using Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance issue.
- **Model Building and Training**
  - We trained various machine learning models and ensembles of different combinations of these classifiers, as well as a Convolutional Neural Network (CNN) and Feed Forward Neural Networks for pose detection. Hyperparameter tuning was conducted for certain models, including using Grid Search for Support Vector Machine (SVM).
- **Real Time Implementation**
  - We implemented our models for real-time pose detection, and achieved promising results. In addition, we developed a web interface for our project, which allows users to upload images for pose detection.

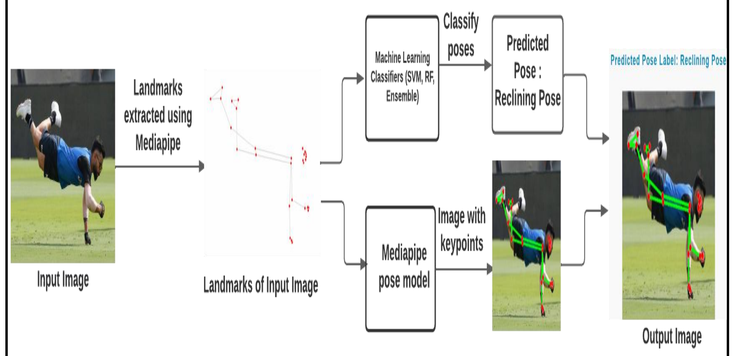

Figure 1: Overview of Proposed Methodology

Experimental Results

- **Performance Evaluation**

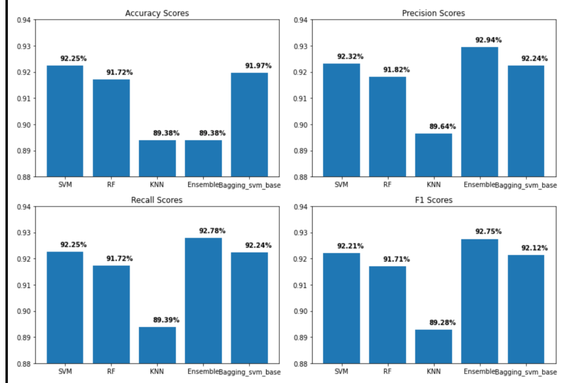

Figure 2: Performance Evaluation of Classification Models after 10 Fold Cross Validation

- Feed Forward Neural Network (Multi Layer Perceptron) and Custom Convolutional Neural Network achieved an accuracy of 89.5% and 89.4% respectively.
- After applying data augmentation techniques on the image dataset, AlexNet model was trained and achieved an accuracy of 69.82% on the test set."

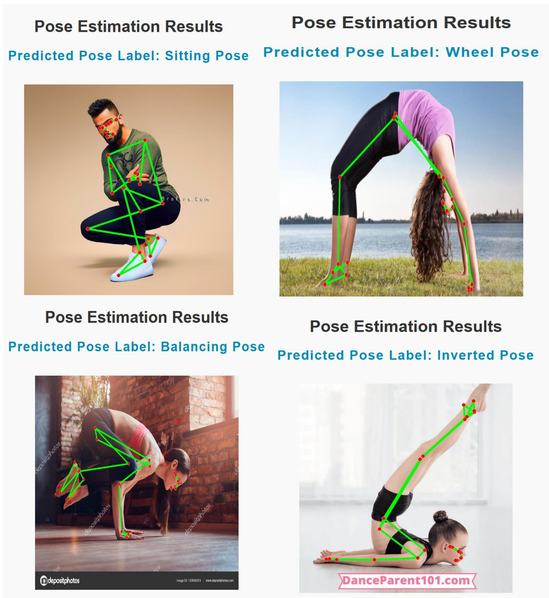

Figure 3: Experimental Results

Conclusions

- A novel machine learning method is proposed, leveraging 3D coordinates (width, height, and depth) for more accurate pose estimation in human pose detection.
- Our method achieved an accuracy of 92% using SVM and Ensemble methods after performing 10-fold Cross Validation, demonstrating the effectiveness of combining Mediapipe for key point extraction and supervised learning for pose detection.
- This approach has significant implications in various fields, including sports and fitness tracking, health monitoring, and human-computer interaction..

Selected References

[1] “Yoga-82: A New Dataset for Fine-grained Classification of Human Poses”, Manisha Verma and Sudhakar Kumawat and Yuta Nakashima and Shanmuganathan Raman, (2020). https://doi.org/10.48550/arXiv.2004.10362

[2] “Yoga Pose Detection and Correction using Posenet and KNN”, Varsha Bhosale, Pranjal Nandeshwar, Abhishek Bale, Janmesh Sankhe, International Research Journal of Engineering and Technology (2022)

[3] “Yoga Pose Classification Using Deep Learning”, Shruti Kothari, Robert Chun, Katerina Potika, and Susmit Gaikwad.