Comparative Analysis of Machine Learning Models for Yoga Pose Prediction using Mediapipe: A Multi-Dataset Study

Abstract:

The surge in yoga's popularity, coupled with the integration of technology into health and fitness practices, has given rise to the development of pose prediction models. This study endeavours to meticulously scrutinize and juxtapose five diverse datasets for yoga pose prediction, utilizing the powerful Mediapipe framework. Furthermore, the research aims to gauge the efficacy of five distinct machine learning models when applied to yoga pose prediction, with a focus on the statistical metrics derived from test data obtained by executing the algorithm on images of poses demonstrated by a trained yoga instructor. The examination encompasses five specific yoga poses: tree pose, downward dog, warrior 1, warrior 2, and upward plank.

Introduction:

1.1 Background:

Yoga has garnered widespread recognition for its multifaceted benefits to physical and mental well-being. The integration of technology into yoga practices has given rise to pose prediction models, offering practitioners a tool to refine their postures. The Mediapipe framework serves as a robust platform, enabling precise pose estimation through the extraction of key points from images or videos.

1.2 Objectives:

Conduct an in-depth analysis and comparison of five distinct datasets for yoga pose prediction.
Evaluate the performance of five machine learning models within the Mediapipe framework.
Identify the most effective model based on comprehensive statistical metrics.
Methodology:

2.1 Datasets:

Carefully select five datasets representing diverse yoga poses, backgrounds, and practitioners, ensuring a comprehensive range of scenarios to enhance model generalization.

Dataset 1: Kaggle-sourced data involving internet-scraped images of five poses.
Dataset 2: Curated dataset derived from Dataset 1, incorporating manual elimination of pose modifications.
Dataset 3: Yoga-82 Dataset.

Dataset 4: Curated dataset derived from Dataset 3, featuring manual elimination of pose modifications.
Dataset 5: Images of the studied poses extracted from popular YouTube videos at specific timestamps.

Trained on Kaggle Train dataset

2.2 Processing:

Utilize Mediapipe landmark detection to extract joint angles from processed images, employing a statistical approach (Z distribution) to establish angle ranges. Generate a CSV file for each pose, detailing various joint angles.

2.3 Real-time Evaluation:

During testing, employ an RGB camera for real-time image processing through Mediapipe, extracting joint angles, and comparing them against predetermined angle ranges. The system predicts pose accuracy or suggests corrections based on identified discrepancies.

Evaluation Metrics:

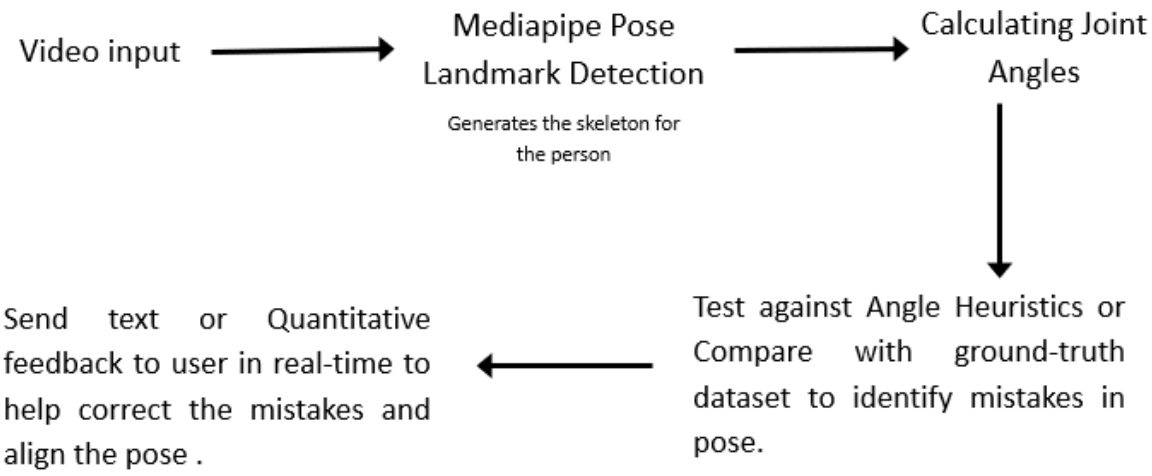
3.1 Accuracy:

Measure the overall accuracy of each model in predicting yoga poses on test datasets against the ground truth established by a study performed with a trained instructor.

Results and Discussion:

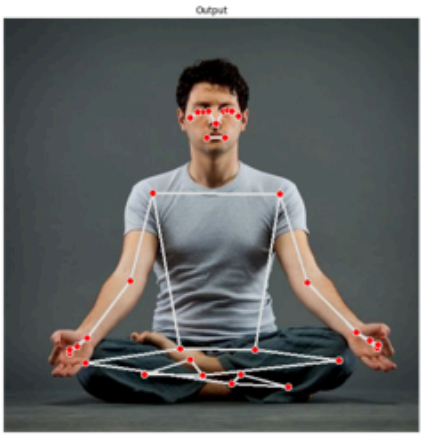
Present the findings of the comparative analysis, delineating the strengths and weaknesses of each model across chosen metrics. Discuss implications, potential applications, and avenues for future research.

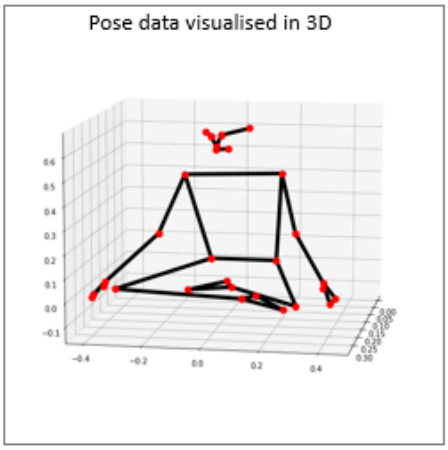
The pipeline:



What is MediaPipe?
MediaPipe is an open-source framework for building pipelines to perform computer vision inference over arbitrary sensory data such as video or audio. Using MediaPipe, such a perception pipeline can be built as a graph of modular components.
What is a computer vision pipeline?
In computer vision pipelines, those components include model inference, media processing algorithms, data transformations, etc. Sensory data such as video streams enter the graph, and perceived descriptions such as object-localization or face-keypoint streams exit the graph.
What is MediaPipe used for?
The MediaPipe framework is mainly used for rapid prototyping of perception pipelines with AI models for inferencing and other reusable components. It also facilitates the deployment of computer vision applications into demos and applications on different hardware platforms. The configuration language and evaluation tools enable teams to incrementally improve computer vision pipelines.
What are the advantages of MediaPipe?
1) End-to-end acceleration: Use common hardware to build-in fast ML inference and video processing, including GPU, CPU, or TPU.
2) Build once, deploy anywhere: The unified framework is suitable for Android, iOS, desktop, edge, cloud, web, and IoT platforms.
3) Ready-to-use solutions: Prebuilt ML solutions demonstrate the full power of the MediaPipe framework.
4) Open source and free: The framework is licensed under Apache 2.0, fully extensible, and customizable

Extracted pose skeleton from Mediapipe:





Perform Landmark Detection : mediapipe is used to create a skeleton of the person performing the yoga poses and present the results as 33 simple key points: