IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 1, January 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201015 |

A Comparative Study of Human Pose Estimation

Vishnu J G [1], Divya S J [2]

U.G. Student, Department of Computer Science and Engineering, PES University, Bangalore, India[1]

Professor, Department of Computer Science and Engineering, PES University, Bangalore, India[2]

ABSTRACT: A comparative study of the human pose estimation helps us understand the choice of operation model to follow while designing an application for a specific purpose. In this paper we discuss about the features of different human pose estimation models and their relevance and usability in different application.

KEYWORDS: Computer Vision, Machine Learning, Pose estimation, Model building, Human pose estimation, Comparative study, Posenet, Mediapipe, Openpose, YOLO.

I. INTRODUCTION

Over the past decade there has been a drastic improvement in the field of computer vision-based image processing ML models. Computer Vision has enabled many forms of real time processing and analysis of images and videos from the real world so that actionable decisions can be taken on time. But the analysis of these images and videos require lot more tools than just computer vision.

This gave rise to image processing and analysis tools for detecting objects, extracting useful data and performing image modifications. Image detection enabled developers to track objects, monitor behaviour and take actions on them as required. This was further developed to an automated low error systems which could handle monitoring and taking actions at the appropriate situations. This was possible with neural networks like CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network).

Among the many object detection applications detecting human movements and poses stood out to be a major player in this domain. Libraries such as Mediapipe and PoseNet were dedicated just for the purpose of human pose detection. These libraries provide a mapping of the complete skeletal structure of the person in frame. With the help of such models, we could easily identify the activities of a person in real time.

Pose estimation can be used for guiding exercises, detecting a violent action in the input video and much more. Different models provide different functionalities, which are applicable to varied use cases. For example, the PoseNet model provides a 17-point skeletal structure, whereas the MediaPipe models provides 33 key human skeletal points. Here we can see that more accuracy can be achieved with the MediaPipe model, at the same time the delay is relatively more is the later.

In this paper we discuss about the various pose estimation libraries, their implementation approaches, their advantages and disadvantage and their use cases. We compare the accuracy, precision and the ability of the model to detect and map the skeleton in different conditions such as low lighting, blurry images and multi-user detection.

II. PROBLEM STATEMENT

This research pertains to a comparative study of different pose estimation models. While building an application based on human pose detection, there are many pose detection libraries to choose from, each of these libraries come with their own set of features. This confuses the developer on the choice of operation model to go ahead with.

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201015 |

Planning the operational model beforehand is necessary in development. This reduces development cost and reduces the risk of hitting a dead-end due to lack of a feature or some other performance issue faced during the development or post- development. To reduce re-development costs, companies and individual developers wish to choose a library which fulfills the requirements and also the performance expectations.

III. COMPARISON PARAMERTERS

The most obvious requirement would be the number of skeletal points in the mapping of the human body. Choosing a model with more points helps in improving accuracy and opens more opportunities for the developers. Whereas choosing a model that performs its functionalities more efficiently while providing limited features, such as lesser skeletal points maybe suitable in systems with limited hardware resources.

There is an expectation from these libraries to pre-process the input images or videos for better detection and analysis. One such condition would be an image taken in a dimly lit room. Some libraries drop this pre-processing assuming that the image will be from a brightly lit room, thereby reducing processing time. But this may not always be desirable by the developers as real-world applications are expected to work even in dimly lit rooms with the claimed accuracy.

We can also see requirements wherein detection of a single user is necessary even in a frame with multiple users visible. Similarly, some applications require all the users in the frame to be detected and the pose to be estimated in near real time. This brings up a lot of performance issues as now we are dealing with multiple skeletal mappings from just a single frame of a video. The features and abilities of these libraries vary based on their implementation procedure followed.

Processing time also matter in the application, real time application requires the processing to be done in real time or near real time. Few applications do not actually require immediate processing instead the application is expected to do the processing on huge volumes of data from a database.

The following are the parameters used to compare the pose estimation models:

1. Single user or Multi-user
2. Speed and Accuracy
3. Hardware Support

IV. THE LIBRARIES AND THEIR IMPLEMENTATIONS

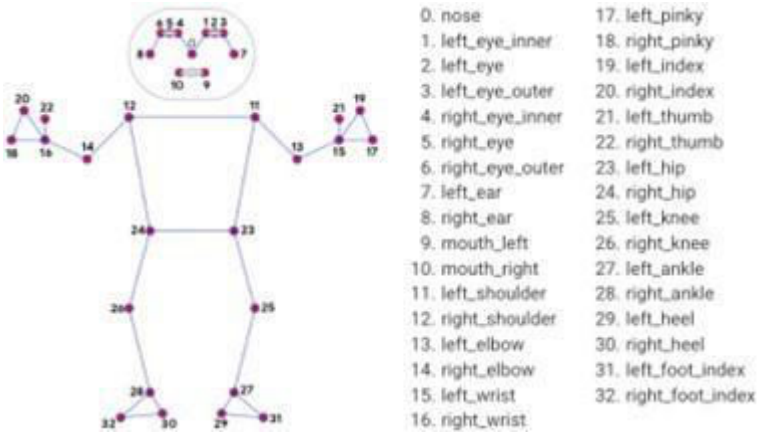
1. MoveNet: Released on 15 May 2021, MoveNet leverages on heatmaps to accurately localize human key points using a feature extractor and a set of prediction heads. The prediction generally follows CenterNet, improving both speed and accuracy with a few key changes. All models are trained using the TensorFlow Object Detection API. MobileNetV2 is used with an attached feature pyramid network (FPN), allowing for a high resolution (output stride 4) and a semantically rich feature map output. A Temporal Filter is tuned to simultaneously suppress high-frequency noise (jitter) and outliers from model while optimizing throughput for quick motions.

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201015 |

2. MediaPipe: MediaPipe offers cross-platform, customizable ML solutions for live and streaming media. It has built-in fast ML inference and processing accelerated even on common hardware, free to use and open source with framework and solutions both under Apache 2.0, fully extensible and customizable. MediaPipe (BlazePose) extended the idea from Stack Hourglass and used an encoder-decoder network architecture to predict heatmaps for all joints, followed by another encoder that regresses directly to the coordinates of all joints. This enables heatmap branch to be discarded during inference so that it is lightweight enough to run on a mobile phone. At 30 frames inference speed with 33 points with multiple versions offering up to 543 key points for pose, face, and hands, it is an ideal choice to consider for cross platforms deployment especially in fitness applications.



3. YOLOv7: The YOLO (You Only Look Once) v7 model is the latest in the family of YOLO models. YOLO models are single stage object detectors. In a YOLO model, image frames are featurized through a backbone. These features are combined and mixed in the neck, and then they are passed along to the head of the network YOLO predicts the locations and classes of objects around which bounding boxes should be drawn. YOLO conducts a post-processing via non-maximum suppression (NMS) to arrive at its final prediction.

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201015 |

4. OpenPose: OpenPose is the first real-time multi-person system to jointly detect human body, hand, facial, and foot key points (up to 135 key points) on single images. It leverages on a bottom-up nonparametric representation of association Part Affinity Fields (PAFs), to "connect" and find body joints on an image, associating them with individual people. The original OpenPose was developed based on VGG pre-trained network using Caffe framework. Cao et al focused-on refining PAF, removing body part confidence map refinement while increasing the network depth, resulting in a faster and more accurate model.

V. RESULTS AND DISCUSSION

A. Single User or Multi-User

Open-source tools like the ones mentioned above serve different purposes, therefore have different implementations. Single user recognition maybe necessary in case such as workout application and user identification where at any instance of time only user needs to be identified and tracked. Whereas application which require identification of multiple users from a video require multi-user identification.

MediaPipe is an open-source system from google and allows single user pose estimation, with the user image being mapped to 33 skeletal points in real time. MoveNet is also a similar tool which is google based and is built by a digital healthcare company called IncludeHealth. OpenPose on the other hand is an open-source real time tool to detected multiple users in a single frame and map their skeletal figures. YOLOv7 is a single-stage, multi-person poses estimation model. YOLOv7 pose is unique, as it deviates from the conventional 2-stage pose estimation algorithms.

MediaPipe	MoveNet	OpenPose	YOLOv7
Single User	Single User	Multiuser	Multiuser

B. Speed and Accuracy

This section talks about how fast the model can identify and map the skeletal structure of the user or users in the frame and how close it can get to the actual pose the user or users are currently in. Accuracy is not always desirable as it compromises performance/speed of the application

MediaPipe performs lot better in terms of accuracy as compared to the other models. In low light intensity images, in different orientations, different distances from the camera and in videos with motion, MediaPipe performs better pose estimation than the rest of the models. However, YOLOv7 performs better when there are occlusions in the input image.

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201015 |

Params/Models	MediaPipe	MoveNet	OpenPose	YOLOv7
GPU Support	No	Yes	Yes	Yes
Speed	Low	High	High	Moderate
Low Light Intensity Detection	Good	Fairly Good	Bad	Fairly Good
Detection in presence of Occlusions	Good	Good	Bad	Bad
Overlapping	Fairly Good	Fairly Good	Good	Bad

C. *Hardware Support*

Computational power required for the application varies based on the library used. Some libraries support development of mobile application which does not require much hardware resources and processing power, whereas other require better hardware and can only be built for computers and other specially designed devices.

	Android	iOS	C++	Python	JS
Face Detection	✓	✓	✓	✓	✓
Face Mesh	✓	✓	✓	✓	✓
Iris	✓	✓	✓		
Hands	✓	✓	✓	✓	✓
Pose	✓	✓	✓	✓	✓
Holistic	✓	✓	✓	✓	✓
Selfie Segmentation	✓	✓	✓	✓	✓
Hair Segmentation	✓		✓		
Object Detection	✓	✓	✓		
Box Tracking	✓	✓	✓		
Instant Motion Tracking	✓				
Objectron	✓		✓	✓	✓
KNIFT	✓				
AutoFlip			✓		
MediaSequence			✓		
YouTube 8M			✓		

Fig. 4. MediaPipe Hardware Support

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201015 |

OpenPose Hardware Requirements:

- Ubuntu (tested on 14 and 16) or Windows (tested on 10). Does not support any other OS, but the community has been able to install it on: CentOS, Windows 7, and Windows 8.
- GPU with at least 1.5 GB available (the nvidia-smi command checks the available GPU memory in Ubuntu).
- CUDA and cuDNN installed.
- At least 2 GB of free RAM memory.
- Highly recommended: A CPU with at least 8 cores.

VI. CONCLUSION

We can see that we either compromise more features to achieve performance and accuracy or compromise performance and accuracy for more features in our application. Developers need to keep in mind the features necessary for their requirement while choosing a model for their application. They should also keep in mind the non-functional aspects of the application they are developing. Being too greedy for either performance or for features in the application results in loss of the other.

Furthermore, the developer can keep in mind the implementations of each of these models and few other which have been missed (cause of similarities with the models mentioned), so that better optimized application design can be formulated. A deeper knowledge of the implemented procedure of the models is beneficial both during development and post development.

REFERENCES

1. https://maureentkt.medium.com/selecting-your-2d-real-time-pose-estimation-models-7d0777bf935f .
2. https://learnopencv.com/yolov7-pose-vs-mediapipe-in-human-pose-estimation/ .
3. https://analyticsindiamag.com/tensorflow-releases-new-3d-pose-detection-model/#:~:text=MediaPipe%20BlazePose%20can%20detect%2033,and% 20can%20detect%20multiple%20poses.
4. https://www.hearai.pl/post/14-openpose/
5. Jo, B., Kim, S. (2022). Comparative analysis of OpenPose, PoseNet, and MoveNet models for pose estimation in mobile devices. Traitement du Signal, Vol. 39, No. 1, pp. 119-124. https://doi.org/10.18280/ts.390111
6. https://blog.ambianic.ai/2021/09/02/movenet-vs-posenet-person-fall-detection.html
7. Zhe Cao Gines Hidalgo Tomas Simon Shih-En Wei Yaser Sheikh. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part A nity Fields". In: arXiv:1812.08008v2 (2019).

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details