

YOGA POSE — ESTIMATION

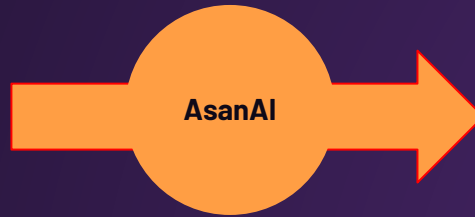


— by

AMISH ASHISH SAXENA
JERRY JOHN THOMAS

— What is it about ?

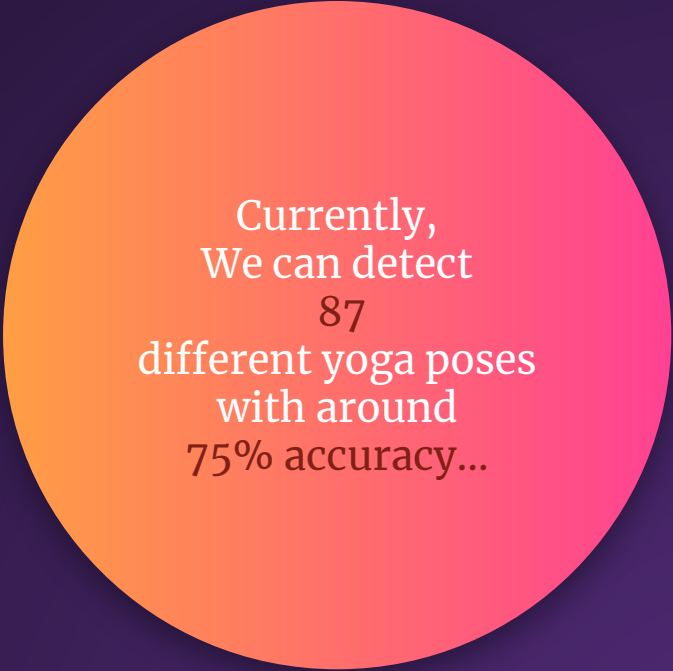
Given a representation of a yoga pose in an image, video or live webcam format, it tells you the yoga pose



PADMASANA

DEMO

- Video (Recorded) demo
- Web based live demo



Currently,
We can detect
87
different yoga poses
with around
75% accuracy...

On Video Labelling Sample



— Web Interface Sample

Realtime Yoga Pose Detection

Created by Amish and Jerry



Your Yoga Pose

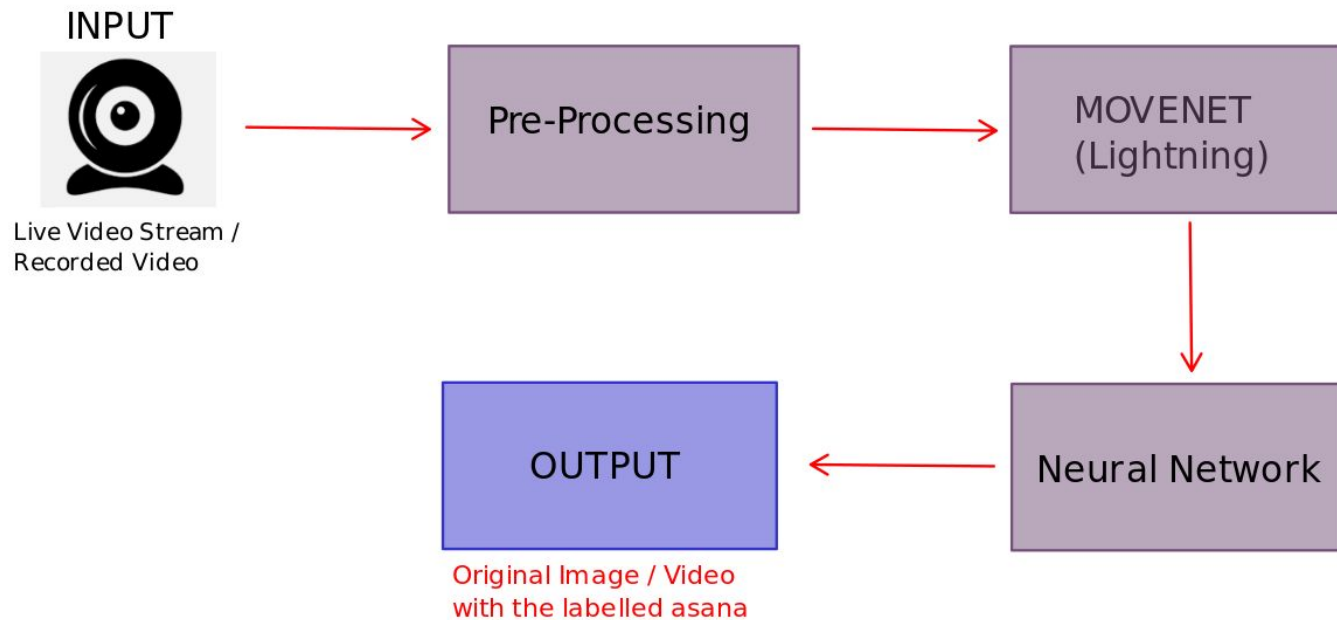
Tadasana

Instructions For Camera Setup:

Stand More Than 3 Feet From The Camera

How ?

HIGH LEVEL DIAGRAM



— Behind the Scenes

Human Pose Estimation

OPENPOSE

MOVENET

ALPHAPOSE

POSENET

Model	Accuracy	GPU Memory Usage	Processing speed (fps)
AlphaPose	low	57.1%	5.21
	default	73.4%	1.15
	high	70.1%	0.69
OpenPose	low	20.1%	19.17
	default	21.3%	18.39
	high	98.1%	3.47

Benchmark on a 1920x1080 video with 902 frames, 30fps

	MacBook Pro 15" 2019. Intel core i9. AMD Radeon Pro Vega 20 Graphics. (FPS)	iPhone 12 (FPS)	Pixel 5 (FPS)	Desktop Intel i9-10900K. Nvidia GTX 1070 GPU. (FPS)
WebGL	104 77	51 43	34 12	87 82
WASM with SIMD + Multithread	42 21	N/A	N/A	71 30

Inference speed of MoveNet across different devices and TF.js backends. The first number in each cell is for Lightning, and the second number is for Thunder.

Movenet Statistics

Google Colab (free)	CPU only Ryzen 7 4600	CPU + GPU Ryzen 5 3550H GTX 1050 (3GB)
43 FPS	53 FPS	59 FPS

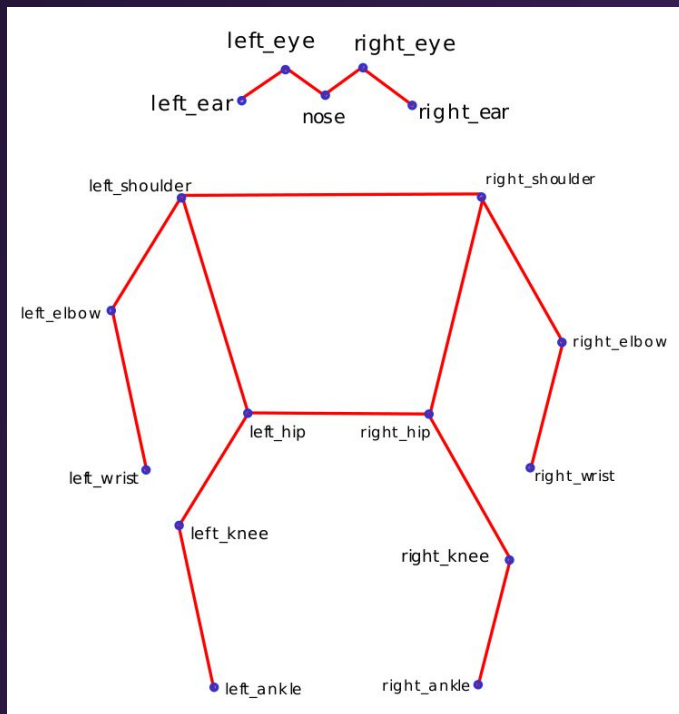
Movenet Lightning Statistics on our experiments

Dataset

- One of the major roadblocks in the project was finding the dataset
- Inspiration : Yoga 82 (old, less functional dataset)
- Solution :
- Web Scrapping + Manual Filtering to get "Yoga 87",

Feature Extraction

17-key points data obtained from the MoveNet



— Feature Extraction

How many features are needed for sufficient accuracy ?

34

46

226

Which features are included ? Why ?

— Feature Extraction

How many features are needed for sufficient accuracy ?

Which features are included ? Why ?

How much accuracy is obtained finally ?

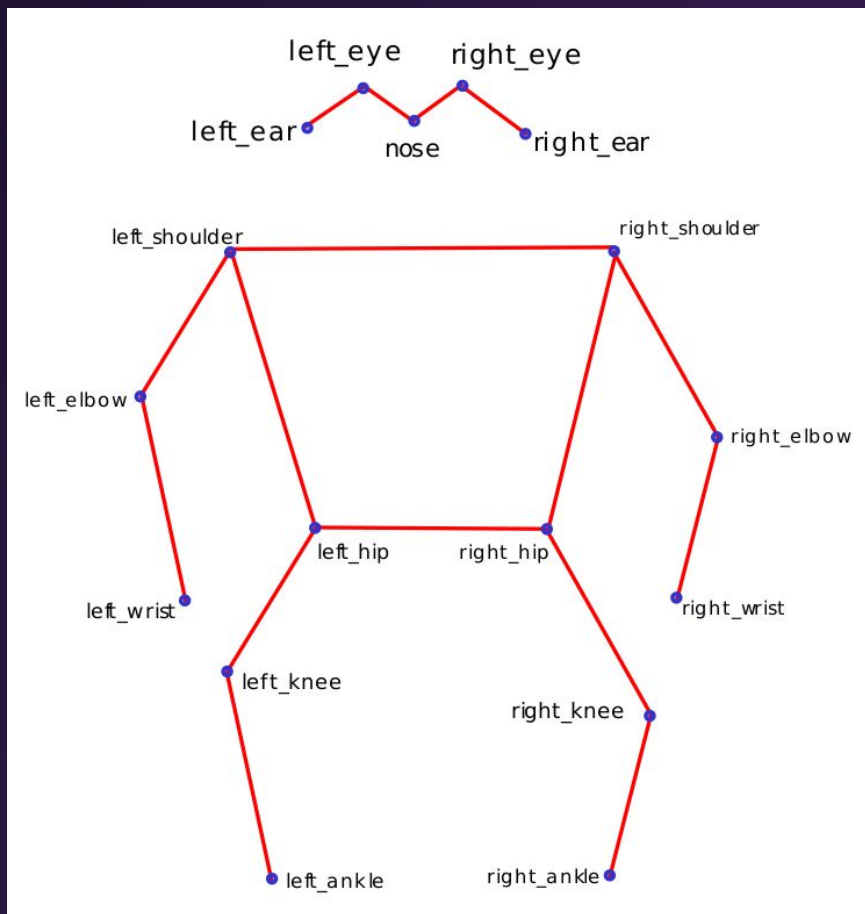
About 75%

34

46

166

226



— Extracted Points

60 points

Booleans representing which among any two pair of key points in higher

34 points

MoveNet (x,y coordinates of 17 points)

12 points

Distance between the left and right - elbows, wrists, knees and ankles

120 points

Euclidean Distance, x-Distance and y-Distance for each of the points to the other one in {elbow, wrist, hip, knee, ankle} ($5C2 * 4 * 3$)

Training

Layer(type)	Output Shape	Param #
dense (Dense)	(None, 256)	58112
dense_1 (Dense)	(None, 256)	65792
dense_2 (Dense)	(None, 128)	32896
dense_3 (Dense)	(None, 87)	11223

Total params: 168,023

Trainable params: 168,023

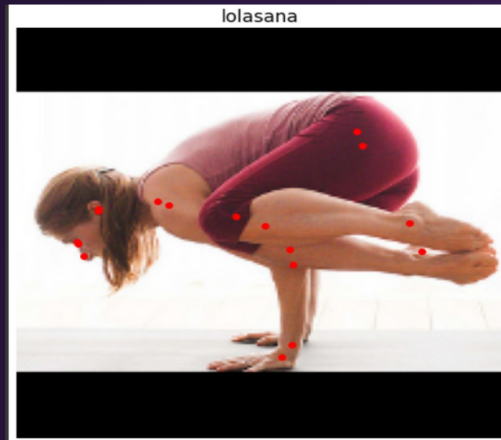
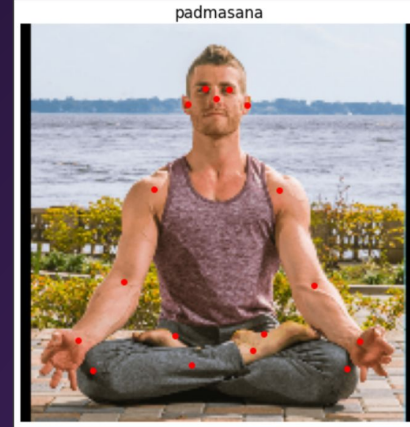
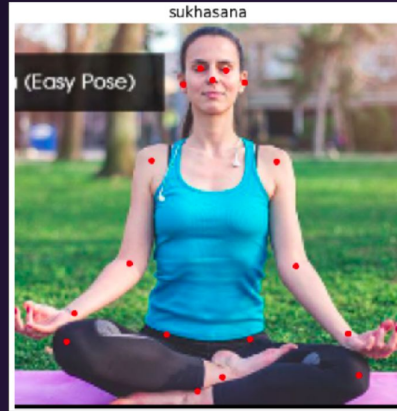
Non-trainable params: 0

— Limitations and Scope for Improvement

- 3d Poses being Recreated in a 2d Photo
- Very Similar Poses
- Camera Orientation

Solution

- Having better and bigger varieties of the dataset from different angles.
- Improving the Model



THANKS!



Any questions?

You can find us at

Amish Ashish Saxena 111901008@smail.iitpkd.ac.in

Jerry John Thomas 111901055@smail.iitpkd.ac.in