MediaPipe: works only for single person

There are ‘solutions’ and ‘tasks’ for pose/hand detection in mediapipe
https://github.com/google/mediapipe/tree/master

Mediapipe ended support for ‘solutions’ on march 1, 2023. Newest versions for detection are released as ‘tasks’. To run the newest models we have to download .task files from overview section of particular models in https://developers.google.com/mediapipe/solutions/guide

‘Solutions’ are giving better results in hand detection than tasks, but solutions are slower as they are not utilizing gpu.
Codes for solutions implementation:
https://github.com/google/mediapipe/tree/master/docs/solutions

In solutions:
Hand detection: Model complexity levels: 0,1
https://github.com/google/mediapipe/blob/master/docs/solutions/hands.md
Pose detection: Model complexity levels: 0,1,2
https://github.com/google/mediapipe/blob/master/docs/solutions/pose.md

In tasks/new solutions:
Hand detection: Model: full
https://developers.google.com/mediapipe/solutions/vision/hand_landmarker
Pose detection: mode: heavy, lite, full
https://developers.google.com/mediapipe/solutions/vision/pose_landmarker

HARDWARE: AMD Radeon Graphics, 4 GB VRAM **integrated**(only 512 mb dedicated), AMD Ryzen 3 5300u processor, 8gb ram, Windows 11

Tested in python 3.11 using mediapipe library for **Tasks**

Benchmarking has been done without drawing the results and without displaying the results. It's only Inference.

Videos: On 1920x1080p video, 25fps original

MODEL: Body pose	**CPU**	**GPU**	**VRAM**	**RAM**	**FPS**	**Processing time per frame**
Heavy	40%	15%	379mb	92%	5	0.16 sec
Lite	59%	30%	371mb	90%	17	0.05 sec
Full	55%	30%	362mb	90%	13	0.07 sec

WebCam: 1080x720p, 30fps: based only on processing without showing output images.

MODEL: Body pose	**CPU**	**GPU**	**VRAM**	**RAM**	**FPS**	**Processing time per frame**
Heavy	60%	28%	354mb	90%	3	0.32 sec
Lite	100%	30%	377mb	86%	8	0.12 sec
Full	100%	34%	400mb	78%	14	0.07 sec

Videos: 3840x2160p, 25fps

MODEL: Hands	**CPU**	**GPU**	**VRAM**	**RAM**	**FPS**	**Processing time per frame**
Full	55%	38%	400mb	90%	11	0.08 sec

WebCam: 1080x720p, 30fps

While running webcam for hand detection using task file, on showing output the code is crashing, this result is based on printing fps without showing image output.

MODEL: Hands	**CPU**	**GPU**	**VRAM**	**RAM**	**FPS**	**Processing time per frame**
Full	41%	15%	366mb	84%	7	0.14 sec

Solutions(prev versions)

In this, cv2.imshow/imwrite is working fine when using webcam, not crashing the code

MODEL:	**CPU**	**GPU**	**VRAM /4gb**	**RAM /7.3gb**	**Cam FPS and processing time per frame**	**Video fps and processing time per frame**
Bodypose: heavy (2)	35%	5-15%	0.5 gb	6.5gb	7fps, 0.14 sec	3fps, 0.25 sec
Bodypose: full (1)	34%	7-19%	0.6 gb	5.8gb	15fps, 0.06sec	6fps, 0.18 sec
Bodypose: lite (0)	36%	8-20%	0.6gb	6.1gb	20fps, 0.05sec	6fps, 0.16sec
Hands: 0	43%	4%	0.7gb	6.1gb	18fps, 0.05sec	2fps, 0.45sec
Hands: 1	40%	3%	0.7gb	6.1gb	15fps, 0.06sec	2fps, 0.43sec

OUTPUTS: (Frames)

Heavy

In heavy, face is detected better than lite and full model
There is a slight more stability in videos generated from heavy, otherwise in lite and full models, landmarks are changing very frequently at some poses.

Stability: heavy>full>lite

Lite: eyes are not detected properly

n li

e

Full model output:

Another heavy model output: fingers are not detected even when visible, lips coordinates are marked, but lips are not visible.

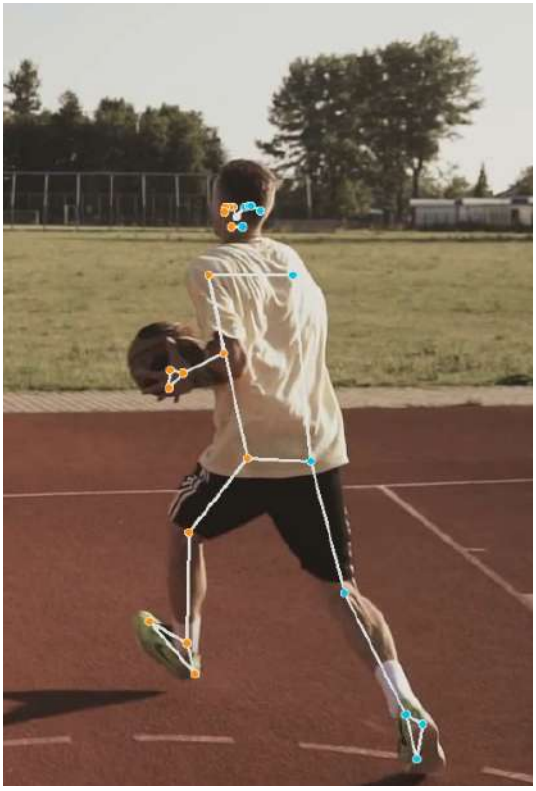
Solutions and tasks comparison

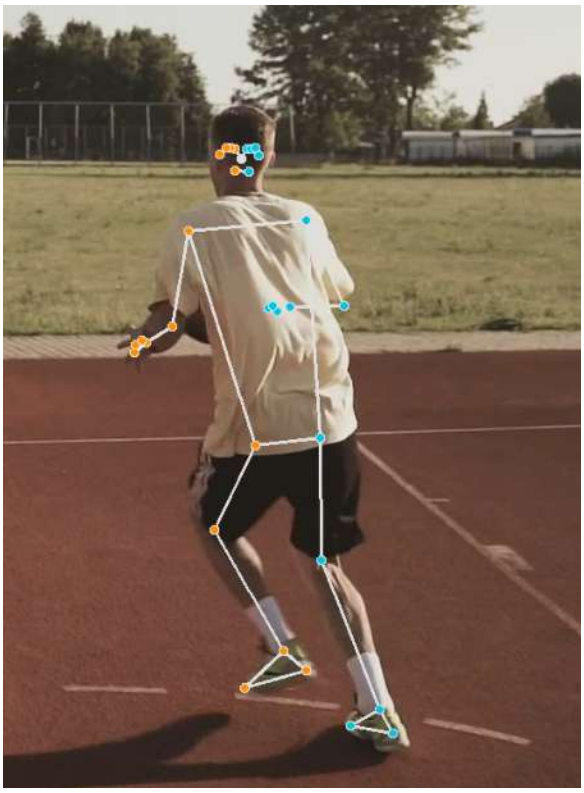
Solutions output for hands: hands detection in every frame

Tasks output on same video: Many undetected frames, on some frames, one hand detection in a few frames.

Solutions output with model complexity:2 **Tasks output with heavy model**

When hand is not visible, newer version of mediapipe pose model is still detecting it, while prev version not.





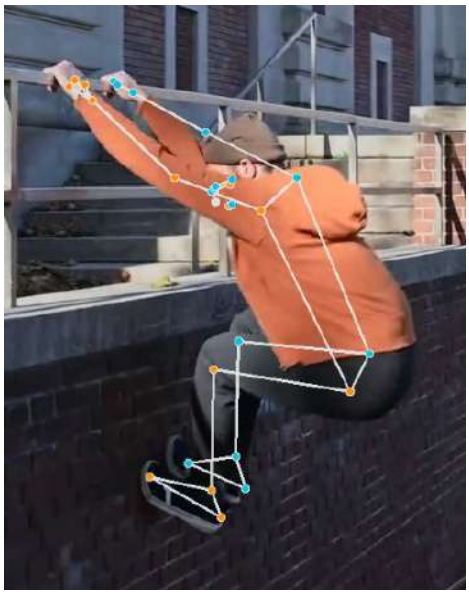
Solutions heavy model has more stability as compared to tasks heavy model, but in this frame, left hand is detected by task's heavy model and undetected in solutions's heavy

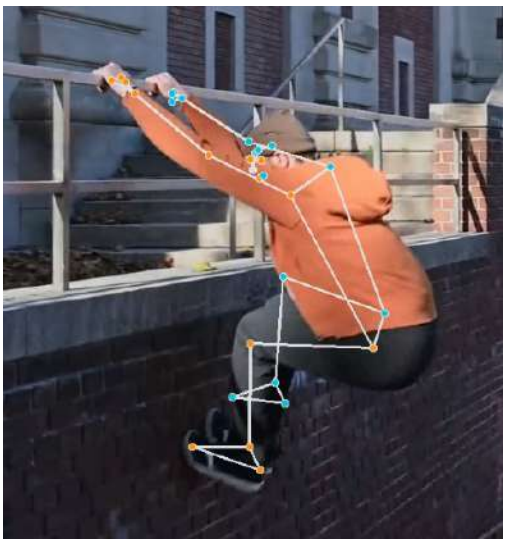
When object is at distant, task's heavy model is detecting it, but not solutions'

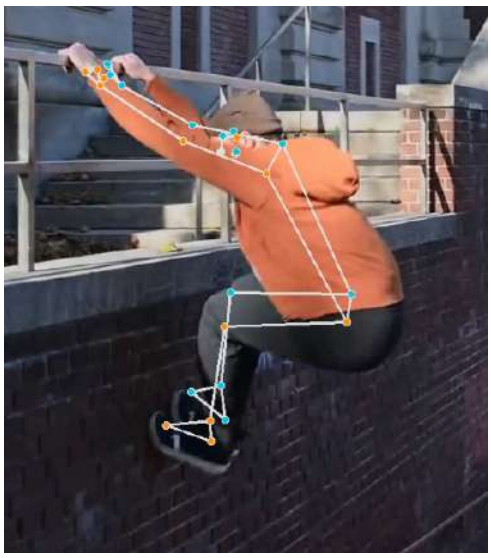
Task:

Heavy Full Lite

Elbow is detected better in full and lite for this frame, leg coordinates are better in heavy and lite

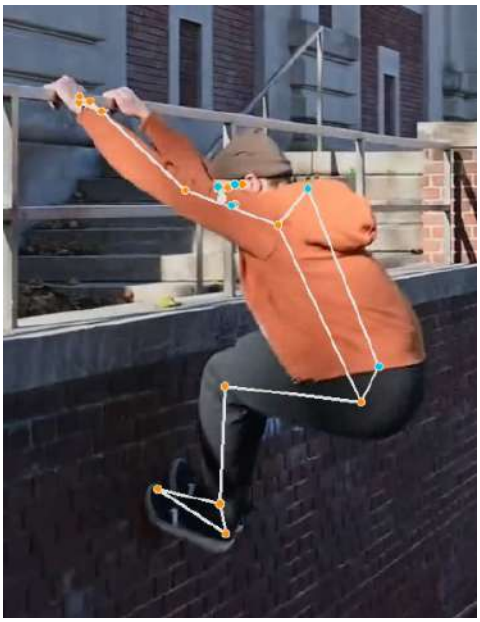
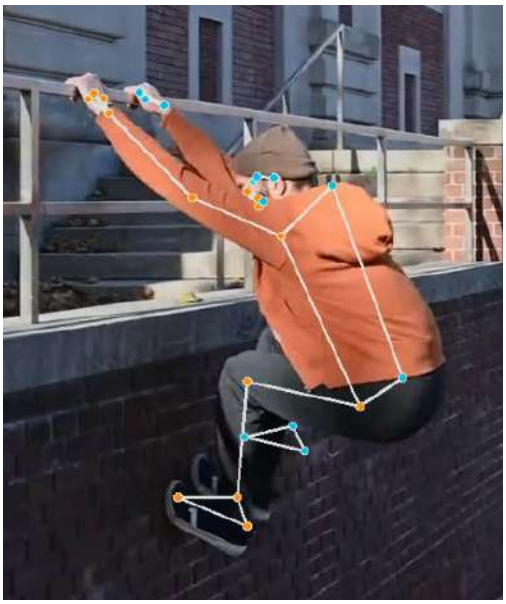
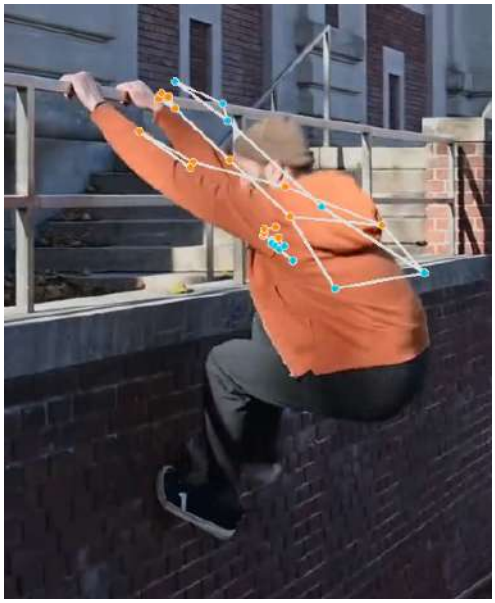






Solutions: when body part is not visible clearly, not detecting it

Heavy **Full** **Lite**

Based on my understanding, in bodypose, newer versions are working better, but in hand detection, there is some issue with newer version as of now, and prev one is working fine.

Performance:
Accuracy:
For body pose detection: task>solutions
In task: accuracy and stability: heavy > full > lite

For hand detection: solutions>task
In solutions: accuracy and stability: model 1 > model 0

Time taken:
For body pose: heavy > full > lite
For hand: only one present in task and in solutions 1>0

GPU Utilisation:
Task models are utilizing gpu, while solutions are not.