Introducing Google AI Edge Portal (https://ai.google.dev/edge/ai-edge-portal): Benchmark Edge AI at scale. Sign-up (https://docs.google.com/forms/d/e/1FAIpQLSfTcGPycQve8TLAsfH46pBlXBZe9FrgJAClwbF7DeL1LgVn4Q/viewform) to request access during private preview.

Pose landmark detection guide

The MediaPipe Pose Landmarker task lets you detect landmarks of human bodies in an image or video. You can use this task to identify key body locations, analyze posture, and categorize movements. This task uses machine learning (ML) models that work with single images or video. The task outputs body pose landmarks in image coordinates and in 3-dimensional world coordinates.

Try it! → (https://mediapipe-studio.webapps.google.com/demo/pose_landmarker)

Get Started

Start using this task by following the implementation guide for your target platform. These platform-specific guides walk you through a basic implementation of this task, including a recommended model, and code example with recommended configuration options:

- **Android** - Code example (https://github.com/google-ai-edge/mediapipe-samples/tree/main/examples/pose_landmarker/android) - Guide (https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker/android)
- **Python** - Code example (https://colab.sandbox.google.com/github/googlesamples/mediapipe/blob/main/examples/pose_landmarker/python/%5BMediaPipe_Python_Tasks%5D_Pose_Landmarker.ipynb) - Guide (https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker/python)
- **Web** - Code example (https://codepen.io/mediapipe-preview/pen/abRLMxN) - Guide (https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker/web_js)

Task details

This section describes the capabilities, inputs, outputs, and configuration options of this task.

Features

- **Input image processing** - Processing includes image rotation, resizing, normalization, and color space conversion.
- **Score threshold** - Filter results based on prediction scores.

Task inputs	Task outputs
The Pose Landmarker accepts an input of one of the following data types: • Still images • Decoded video frames • Live video feed	The Pose Landmarker outputs the following results: • Pose landmarks in normalized image coordinates • Pose landmarks in world coordinates • Optional: a segmentation mask for the pose.

Configurations options

This task has the following configuration options:

Option Name	Description	Value Range	Default Value
`running_mode`	Sets the running mode for the task. There are three modes: IMAGE: The mode for single image inputs. VIDEO: The mode for decoded frames of a video. LIVE_STREAM: The mode for a livestream of input data, such as from a camera. In this mode, resultListener must be called to set up a listener to receive results asynchronously.	{`IMAGE`, `VIDEO`, `LIVE_STREAM`}	`IMAGE`
`num_poses`	The maximum number of poses that can be detected by the Pose Landmarker.	`Integer > 0`	`1`
`min_pose_detection_confidence`	The minimum confidence score for the pose detection to be considered successful.	`Float [0.0, 1.0]`	`0.5`

Option Name	Description	Value Range	Default Value
`min_pose_presence_confidence`	The minimum confidence score of pose presence score in the pose landmark detection.	`Float [0.0, 1.0]`	`0.5`
`min_tracking_confidence`	The minimum confidence score for the pose tracking to be considered successful.	`Float [0.0, 1.0]`	`0.5`
`output_segmentation_masks`	Whether Pose Landmarker outputs a segmentation mask for the detected pose.	`Boolean`	`False`
`result_callback`	Sets the result listener to receive the landmarker results asynchronously when Pose Landmarker is in the live stream mode. Can only be used when running mode is set to `LIVE_STREAM`	`Result Listener`	`N/A`

Models

The Pose Landmarker uses a series of models to predict pose landmarks. The first model detects the presence of human bodies within an image frame, and the second model locates landmarks on the bodies.

The following models are packaged together into a downloadable model bundle:

- **Pose detection model**: detects the presence of bodies with a few key pose landmarks.
- **Pose landmarker model**: adds a complete mapping of the pose. The model outputs an estimate of 33 3-dimensional pose landmarks.

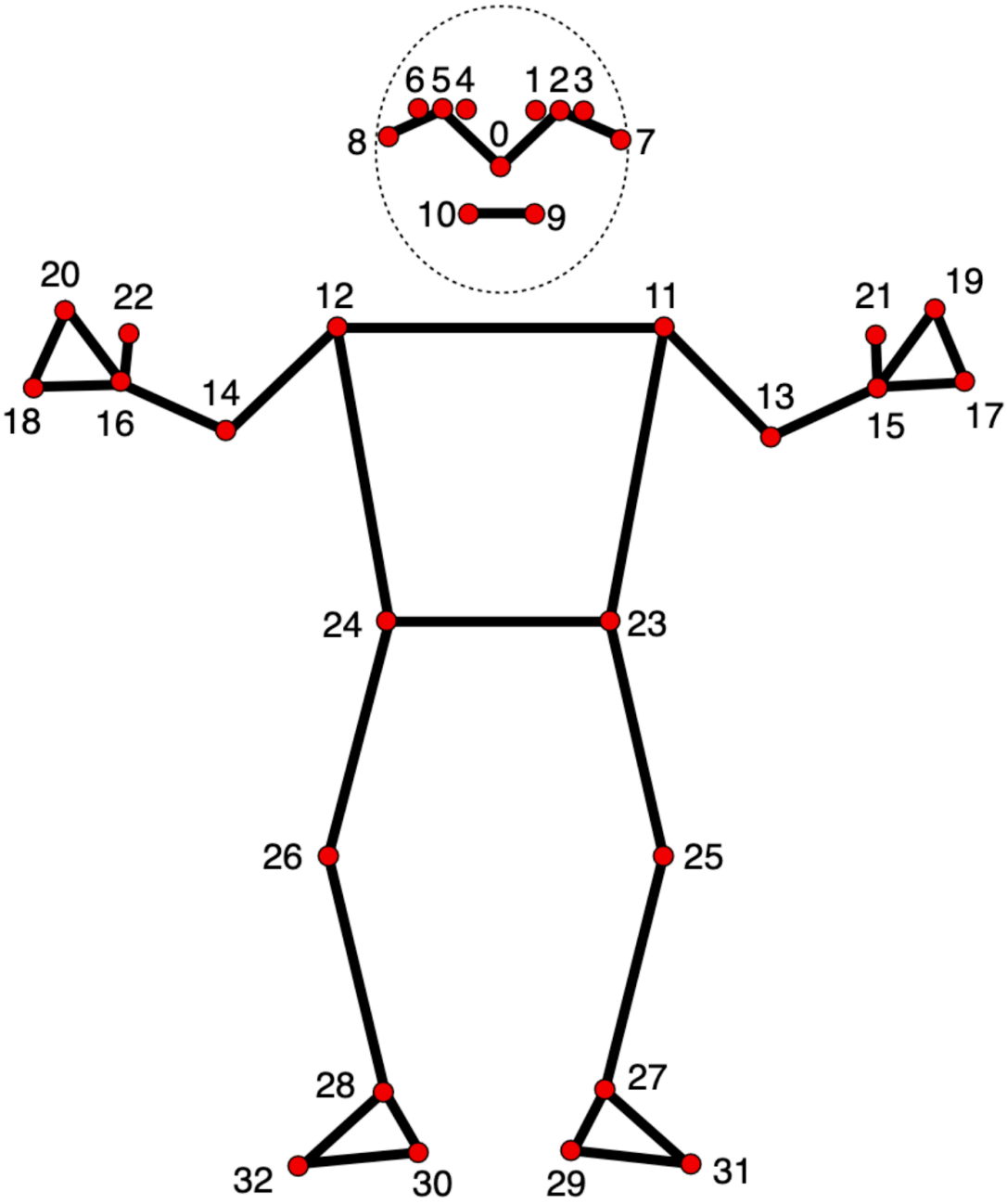
This bundle uses a convolutional neural network similar to MobileNetV2 (https://arxiv.org/pdf/1801.04381.pdf) and is optimized for on-device, real-time fitness applications. This variant of the BlazePose (https://arxiv.org/pdf/2006.10204.pdf) model uses GHUM (https://openaccess.thecvf.com/content_CVPR_2020/papers/Xu_GHUM__GHUML_Generative_3D_Human_Shape_and_Articulated_Pose_CVPR_2020_paper.pdf) , a 3D human shape modeling pipeline, to estimate the full 3D body pose of an individual in images or videos.

Attention: This MediaPipe Solutions Preview is an early release. Learn more (/edge/mediapipe/solutions/about#notice).

Model bundle	Input shape	Dat typ
Pose landmarker (lite) (https://storage.googleapis.com/mediapipe-models/pose_landmarker/pose_landmarker_lite/float16/latest/pose_landmarker_lite.task)	Pose detector: 224 x 224 x 3 Pose landmarker: 256 x 256 x 3	floa 16
Pose landmarker (Full) (https://storage.googleapis.com/mediapipe-models/pose_landmarker/pose_landmarker_full/float16/latest/pose_landmarker_full.task)	Pose detector: 224 x 224 x 3 Pose landmarker: 256 x 256 x 3	floa 16
Pose landmarker (Heavy) (https://storage.googleapis.com/mediapipe-models/pose_landmarker/pose_landmarker_heavy/float16/latest/pose_landmarker_heavy.task)	Pose detector: 224 x 224 x 3 Pose landmarker: 256 x 256 x 3	floa 16

Pose landmarker model

The pose landmarker model tracks 33 body landmark locations, representing the approximate location of the following body parts:



0 - nose
1 - left eye (inner)
2 - left eye
3 - left eye (outer)
4 - right eye (inner)
5 - right eye
6 - right eye (outer)
7 - left ear
8 - right ear
9 - mouth (left)
10 - mouth (right)
11 - left shoulder
12 - right shoulder
13 - left elbow

```
14 - right elbow
15 - left wrist
16 - right wrist
17 - left pinky
18 - right pinky
19 - left index
20 - right index
21 - left thumb
22 - right thumb
23 - left hip
24 - right hip
25 - left knee
26 - right knee
27 - left ankle
28 - right ankle
29 - left heel
30 - right heel
31 - left foot index
32 - right foot index
```

The model output contains both normalized coordinates (`Landmarks`) and world coordinates (`WorldLandmarks`) for each landmark.

Last updated 2025-01-13 UTC.