Ministry of Science and Higher Education of the Republic of Kazakhstan
L.N. Gumilyov Eurasian National University

Faculty of Information Technology
Department of Information Systems

Mediapipe, theory, application, example

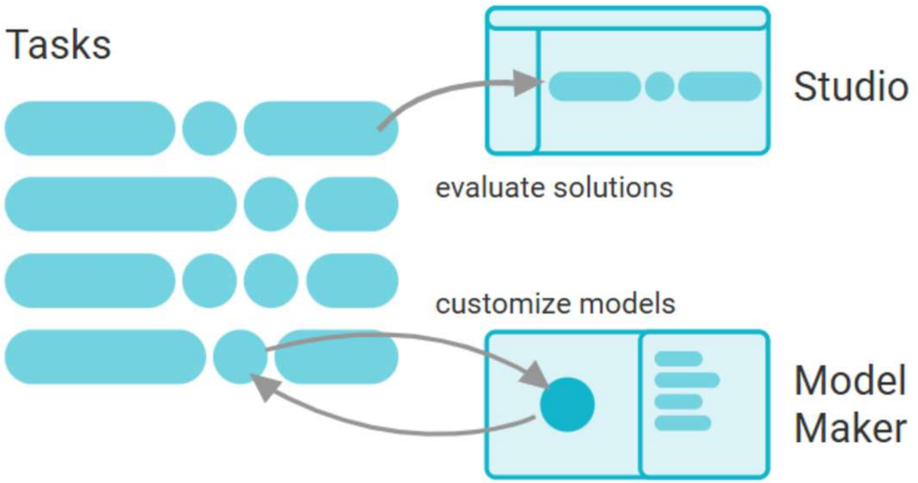
Done by: Iskakov Yerassyl

CHECKED BY: PROF. T.K. ZHUKABAYEVA

INTRODUCTION

MediaPipe is a Framework for building machine learning pipelines for processing time-series data like video, audio, etc. This cross-platform Framework works on Desktop/Server, Android, iOS, and embedded devices like Raspberry Pi and Jetson Nano.

MediaPipe



BRIEF HISTORY

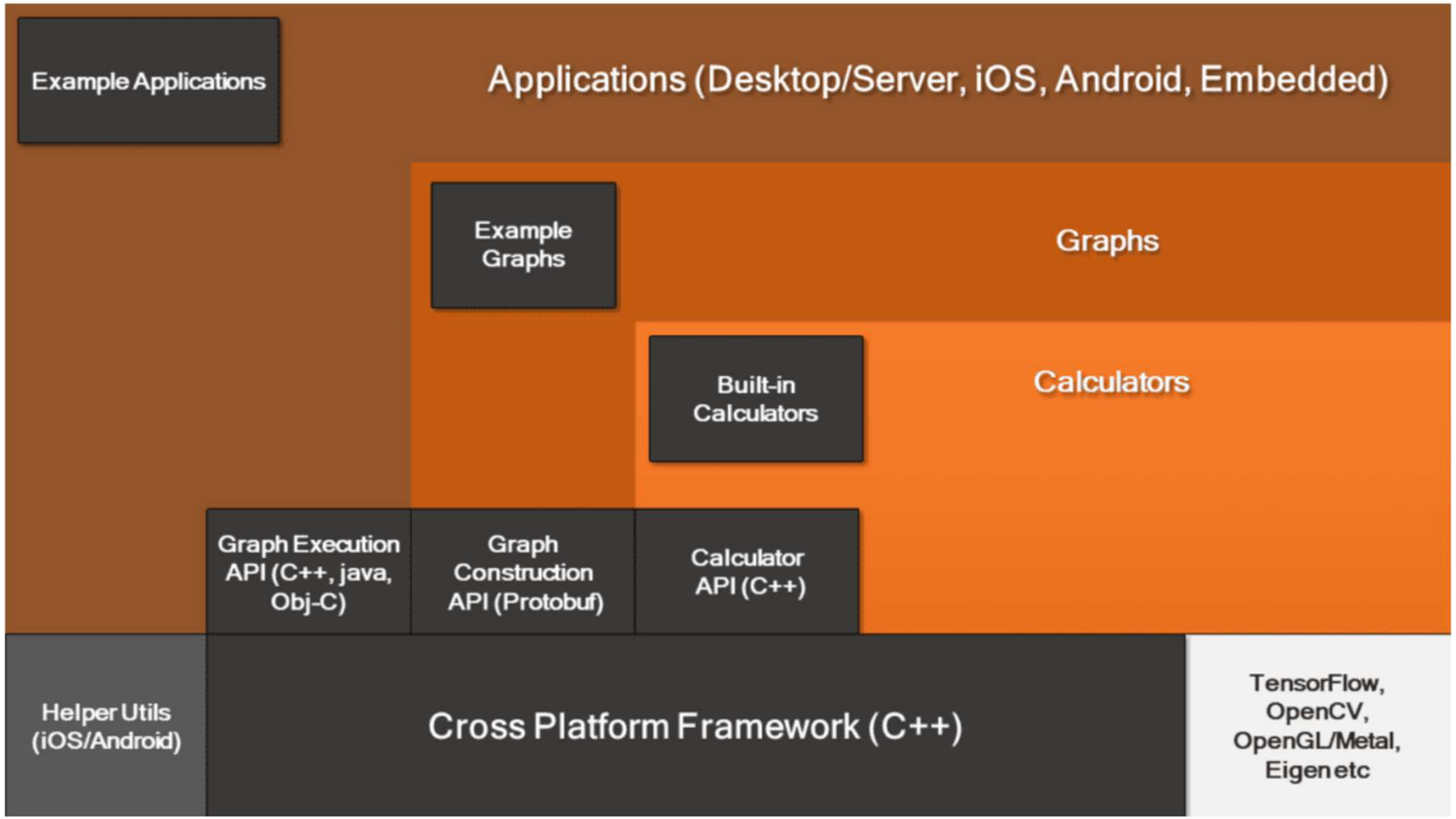
MediaPipe was initially developed by Google in 2012 for real-time video and audio analysis on YouTube. Over time, it was integrated into various Google products and services, including:

- NestCam – Intelligent perception system
- Google Lens – Object detection and recognition
- Google Photos – AI-driven image enhancements
- Google Home – Smart assistant functionalities
- Gmail – Automated email features
- Cloud Vision API – Advanced image analysis

Unlike computationally intensive machine learning frameworks, MediaPipe is optimized for efficiency and requires minimal resources, making it suitable even for embedded IoT devices. In 2019, Google publicly released MediaPipe, enabling researchers and developers to explore new frontiers in real-time perception and AI applications.

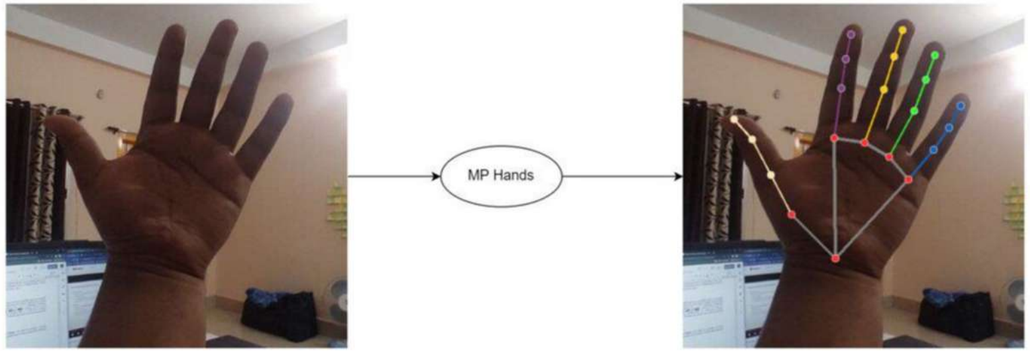
MEDIAPIPE TOOLKIT

MediaPipe Toolkit comprises the Framework and the Solutions. The following diagram shows the components of the MediaPipe Toolkit.



GRAPHS

The MediaPipe perception pipeline is called a Graph. Let us take the example of the first solution, Hands. We feed a stream of images as input which comes out with hand landmarks rendered on the images.



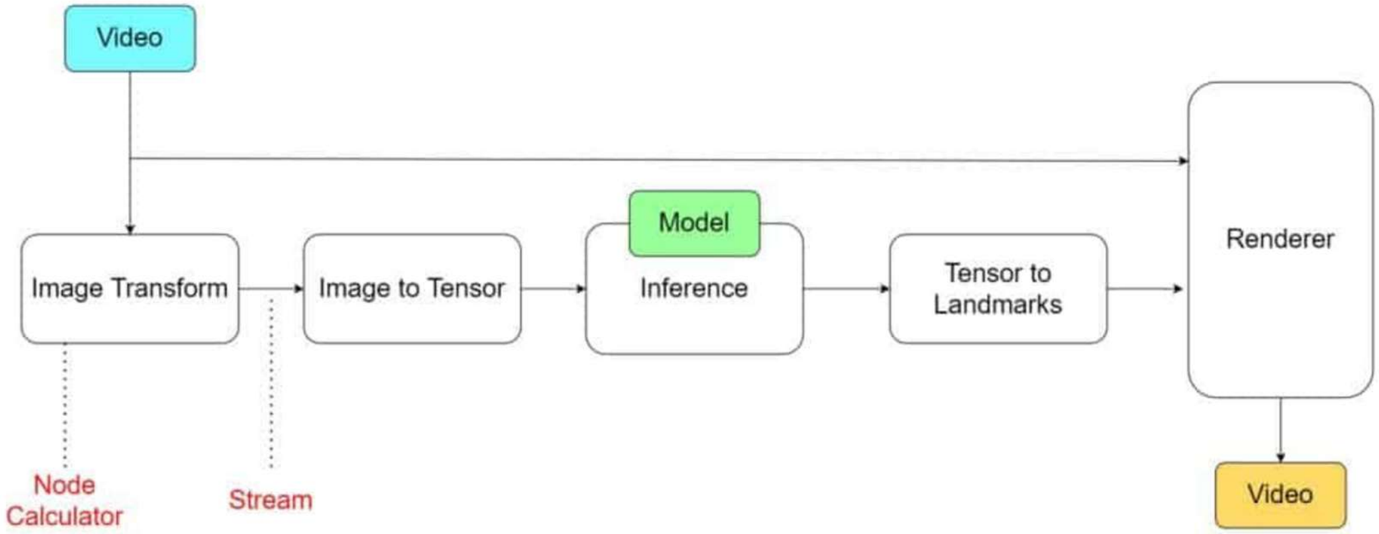


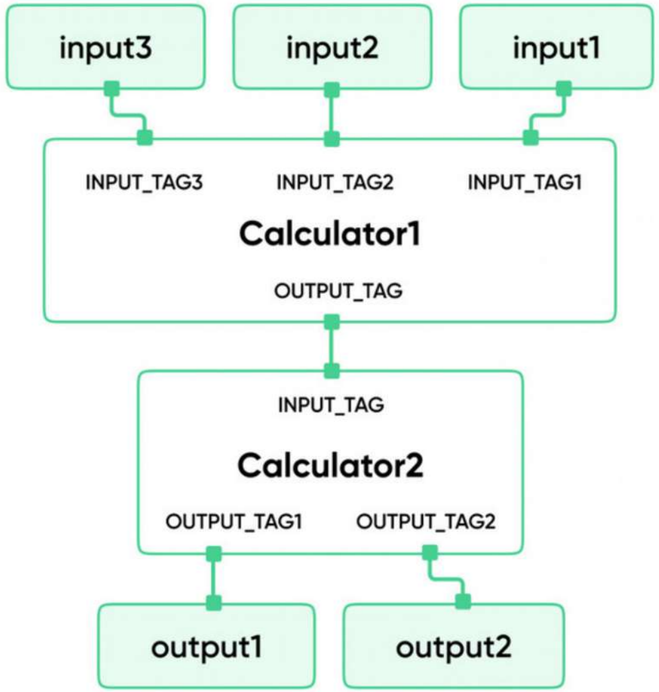
Fig: MediaPipe Hands Solution Graph

In computer science jargon, a graph consists of Nodes connected by Edges. Inside the MediaPipe Graph, the nodes are called Calculators, and the edges are called Streams. Every stream carries a sequence of Packets that have ascending time stamps.

MEDIAPIPE CALCULATORS

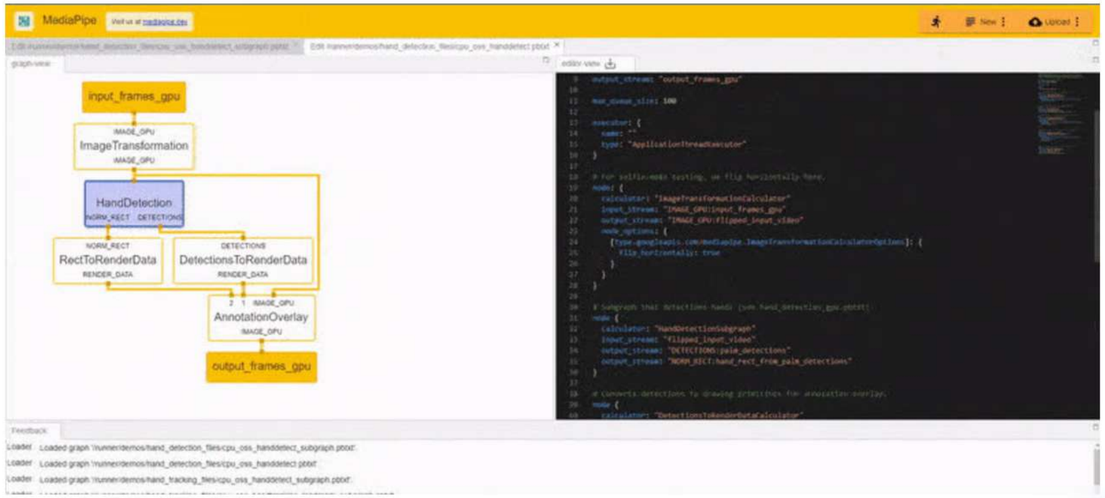
MediaPipe calculators are specialized computation units written in C++, designed to process data packets such as video frames or audio segments. These packets enter and exit through designated ports, enabling structured data flow within a processing graph. Each calculator follows a standard execution lifecycle:

- Open() – Initializes the calculator and declares the payload type.
- Process() – Executes whenever a data packet enters, performing the assigned computation.
- Close() – Finalizes operations when the graph completes processing.



Calculator Types in MediaPipe

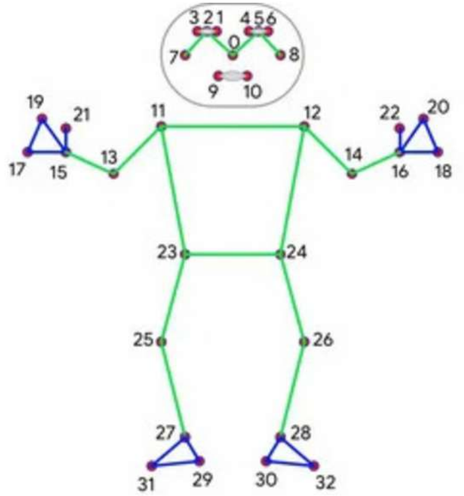
- Pre-processing Calculators – Handle media transformation tasks (e.g., ImageTransform, ImageToTensor).
- Inference Calculators – Enable native integration with TensorFlow and TensorFlow Lite for machine learning inference.
- Post-processing Calculators – Perform ML tasks such as detection, segmentation, and classification (e.g., TensorToLandmark).
- Utility Calculators – Provide additional functionalities like image annotation.



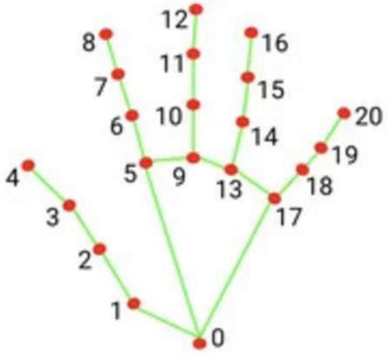
APPLICATION THEORY

Today, there are multiple interesting real-world applications of MediaPipe, in Computer Vision.

- The program uses MediaPipe Tracking to detect a user's body part and track key landmarks, specifically the thumb and index finger.
- The calculated distance is mapped using NumPy's interpolation function.

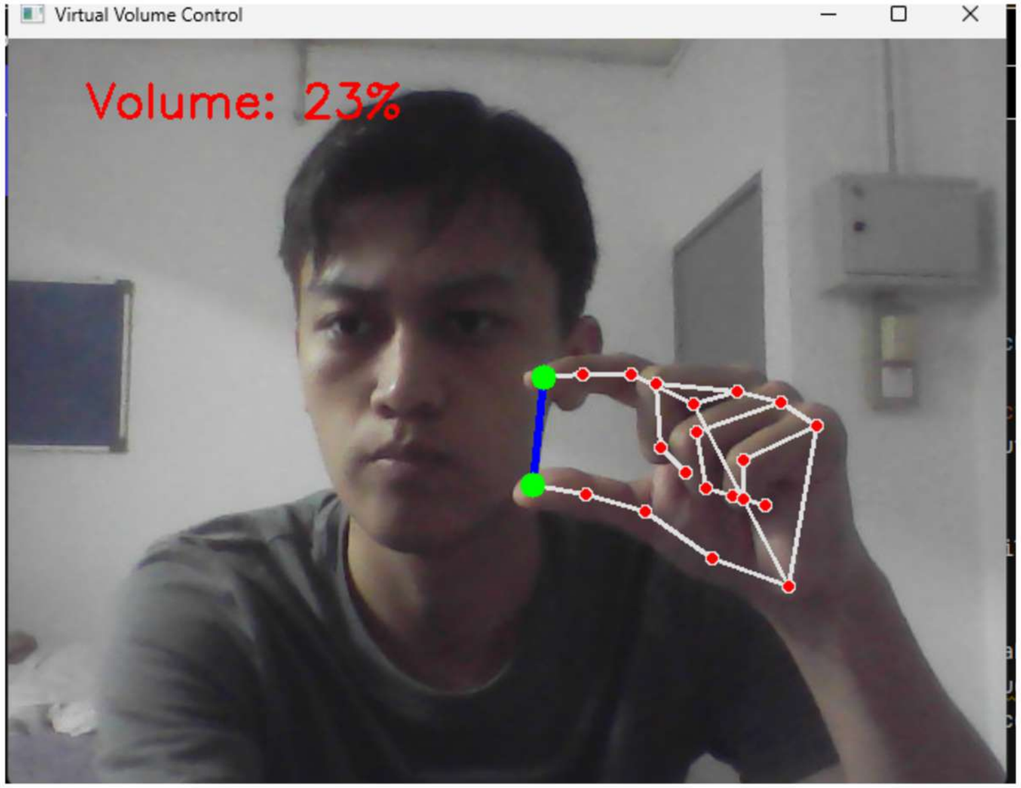


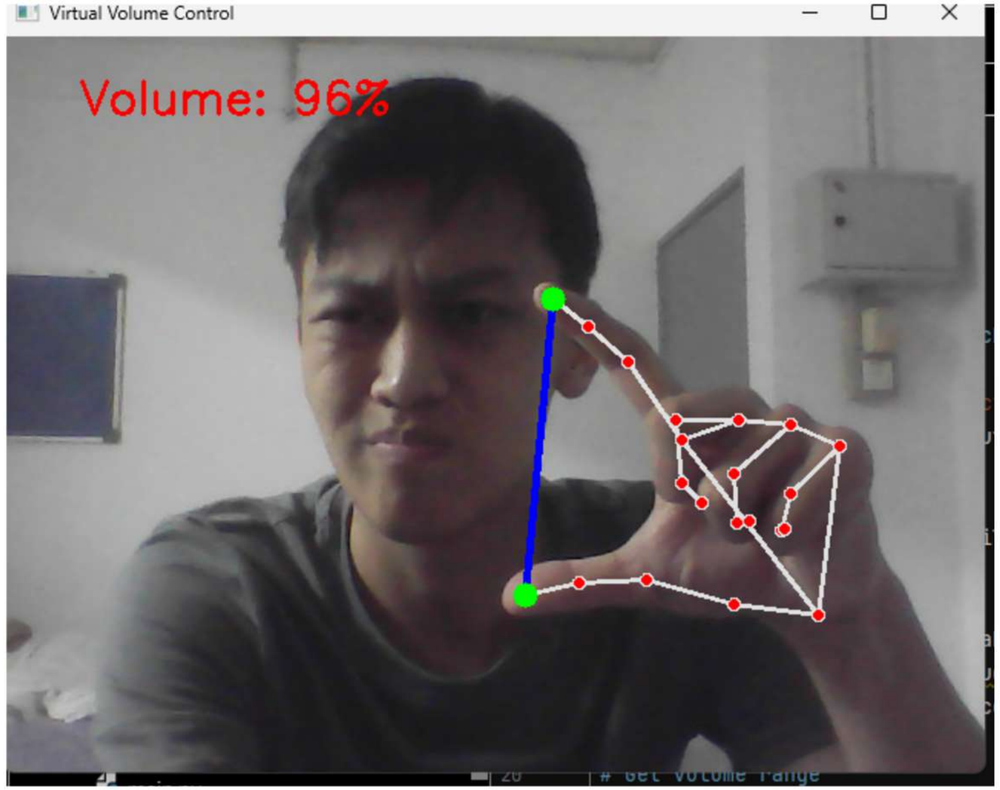
0. nose
1. right_eye_inner
2. right_eye
3. right_eye_outer
4. left_eye_inner
5. left_eye
6. left_eye_outer
7. right_ear
8. left_ear
9. mouth_right
10. mouth_left
11. right_shoulder
12. left_shoulder
13. right_elbow
14. left_elbow
15. right_wrist
16. left_wrist
17. right_pinky_1
18. left_pinky_1
19. right_index_1
20. left_index_1
21. right_thumb_2
22. left_thumb_2
23. right_hip
24. left_hip
25. right_knee
26. left_knee
27. right_ankle
28. left_ankle
29. right_heel
30. left_heel
31. right_foot_index
32. left_foot_index



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

EXPERIMENTAL RESULTS





I have implemented a virtual volume control using hand gestures. The user can adjust the system volume by moving their fingers closer or farther apart.

MEDIAPIPE COMPARISON WITH OTHER OBJECT DETECTION ALGROTHMS (MY OPINION)

- MediaPipe is optimized for real-time, whereas other object detection algorithms like YOLO, Faster R-CNN, require **higher computational resources.**
- Unlike YOLO and Faster R-CNN, which focus on general object detection, **MediaPipe specializes in pose estimation, hand tracking, and facial landmark detection**.
- MediaPipe also runs efficiently on CPU and mobile devices, while YOLO and SSD often require GPUs for optimal performance.
- While Faster R-CNN and Yolo provides high accuracy, it is slower compared to MediaPipe's graph-based pipeline.
- MediaPipe is very easy to implement with built-in solutions, while other algorithms may require custom training (for example CNN based) and tuning for specific object detection tasks.
- Unlike traditional object detection models that require extensive dataset training, MediaPipe **offers pre-trained solutions.**
- **Interesting fact that** MediaPipe's open-source framework provides cross-platform support, making it accessible for developers working on web, mobile, and embedded systems.

https://ai.google.dev/edge/mediapipe/solutions/guide?hl=ru

CONCLUSION

When you need fastest real-time object detection algorithm

MediaPipe

Others

Mediapipe is a powerful tool for real-time machine learning solutions. It is offering efficient and lightweight processing across multiple platforms. Architecture enables easy integration of computer vision and deep learning models for applications face detection, hand tracking, and object recognition. Mediapipe ensures high-speed performance, making it ideal for real-time applications. The framework's open-source nature fosters continuous improvements and community-driven advancements. Its ease of use, combined with pre-trained models, reduces development complexity. Overall, Mediapipe is a versatile and scalable solution for real-time AI applications.

REFERENCES

1. https://learnopencv.com/introduction-to-mediapipe/
2. https://habr.com/ru/companies/agima/articles/696836/
3. https://habr.com/ru/companies/oleg-bunin/articles/735024/

THANKS FOR ATTENTION!