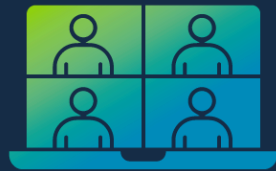


arm AI



Getting started with Arm NN on Android , in just 5 minutes



arm

Welcome!

Tweet us: [@ArmSoftwareDev](https://twitter.com/ArmSoftwareDev) -> #AIVTT

Check out our Arm Software Developers YouTube [channel](#)

Signup now for our next AI Virtual Tech Talk: developer.arm.com/techtalks

Our upcoming Arm AI Tech Talks

Date	Title	Host
November 30 th	Getting started with Arm NN on Android, in just 5 minutes	Arm
December 14 th	Improve PyTorch App Performance with Android NNAPI Support	Arm
January 25, 2022	Peaks, Valleys and Thresholds: The art of segmenting real-time sensor data for tinyML classification, regression and anomaly detection	Reality AI
February 8 th , 2022	Faster time-to-production for computer vision AI on Arm powered edge devices, but just how fast is fast?	Deeplite
February 22, 2022	* New Arm ML Quarterly Research Special * Federated Learning Based on Dynamic Regularization to Debias Model Updates	Arm ML Research

Visit: developer.arm.com/techtalks

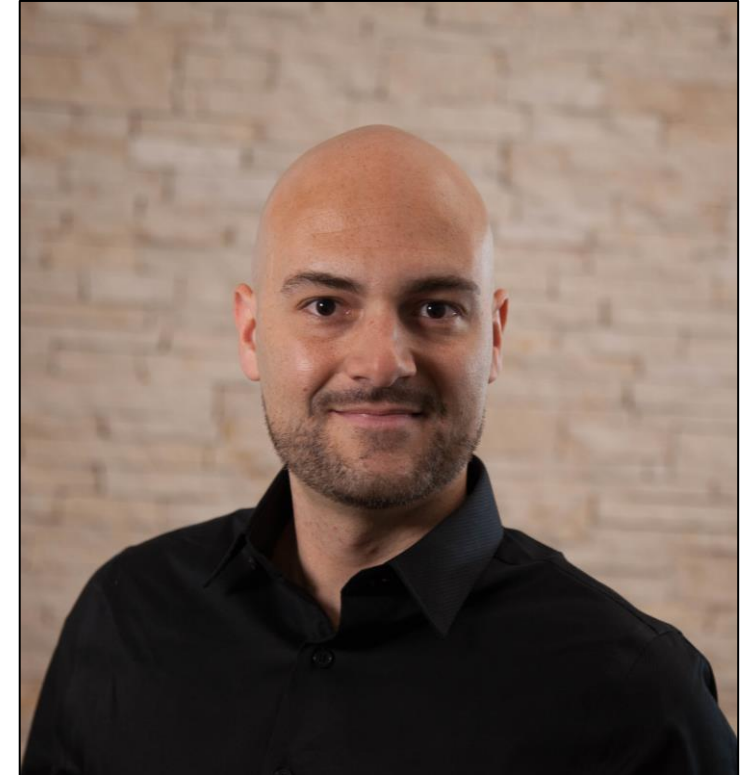
Presenters



Ronan Naughton
Senior Product Manager – ML,
Arm



Jan Eilers
Arm NN Software Engineer,
Arm



Gian Marco Iodice
ACL Staff Software Engineer,
Arm

Introduction - Arm NN and Arm Compute Library

Large Install Base:

- Powering ML in billions of devices:
 - Mobile
 - DTV
 - Automotive
 - Laptop

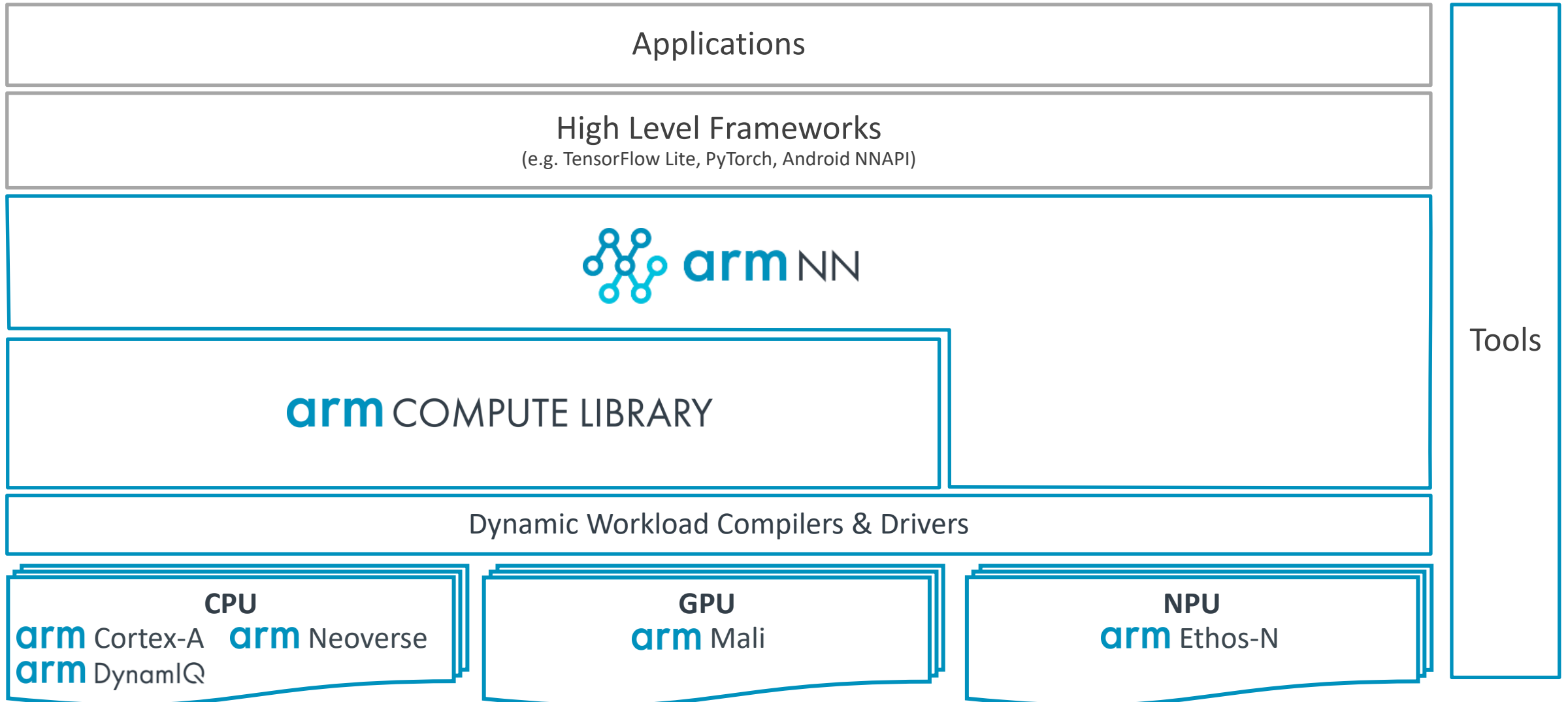
Superior Performance:

- Full Heterogenous execution (CPU, GPU) using a single interface
- Uses advanced network optimization techniques, workload tuning and GEMM heuristics

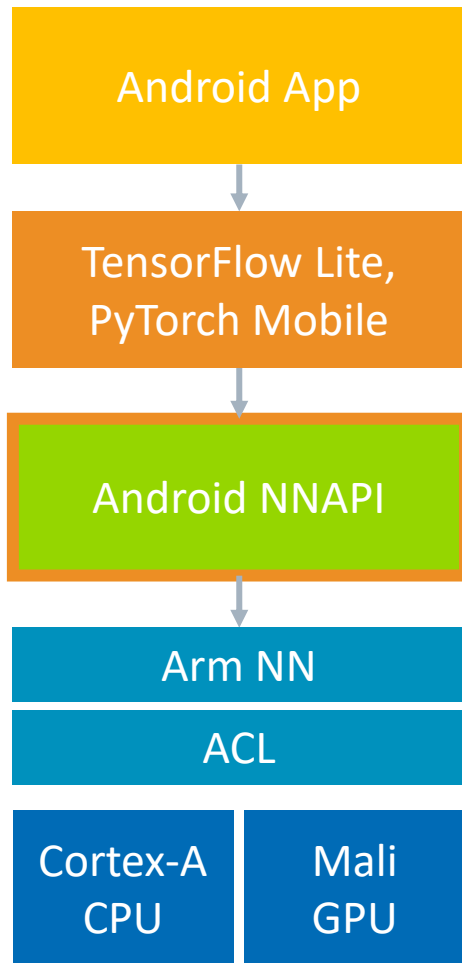
Arm Specific Optimizations:

- Outperforms generic ML libraries due to Arm specific optimization (e.g. dot product for Armv8.2A)
- Quick adoption of new Arm technologies e.g. Armv9-A features

Heterogeneous ML Software and Tools Platform

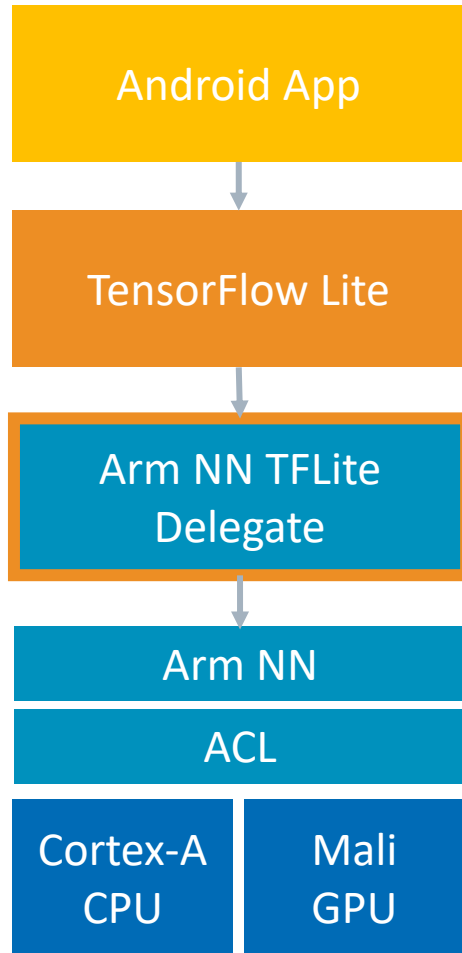


Arm NN on Android – using Android NNAPI



- Pre-installed by vendor on many Android devices
- Support for TensorFlow Lite and PyTorch networks
- Full CTS / VTS compliance
- 72 Operators accelerated with Arm NN backends
- NNAPI model caching supported
- Arm NN HAL driver:
 - `android.hardware.neuralnetworks` HAL interfaces (1.0 – 1.3)

Arm NN on Android - using TensorFlow Lite Delegate



- All TensorFlow Lite Networks can be supported:
 - Key ops (70) – accelerated
 - Unsupported ops – fall back to TensorFlow Lite implementation
- Quick setup time with Android Library (AAR)
- Configurable Optimizations
 - Backend selection
 - fast math
 - fp32->fp16
 - num-of-threads
 - memory-import
- Flexible build options (Cmake, Bazel)

Android Sample applications

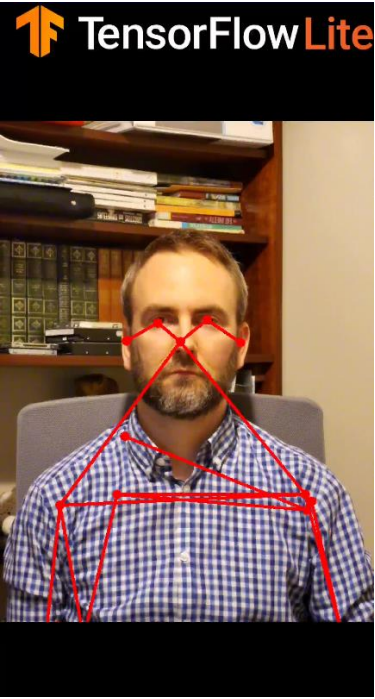
<https://www.tensorflow.org/lite/examples>



Image Segmentation

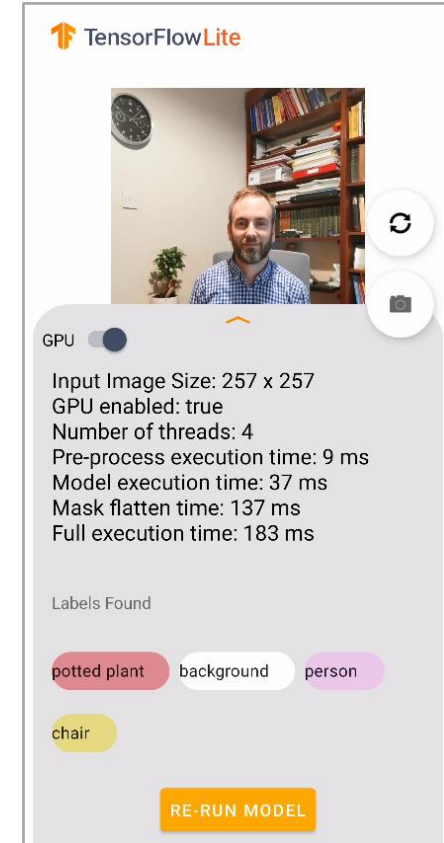
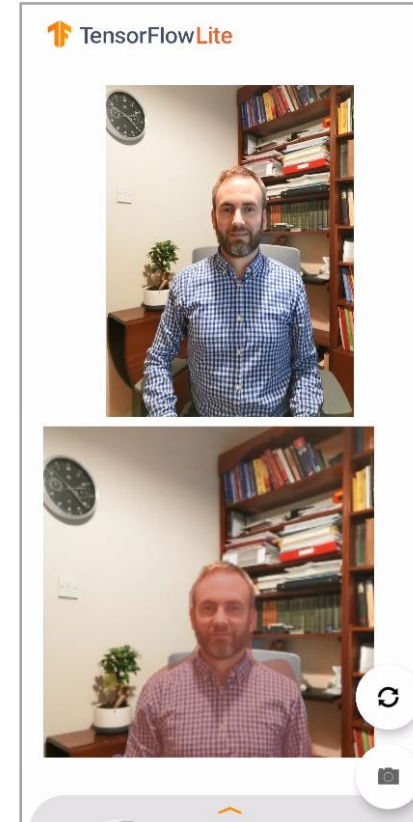
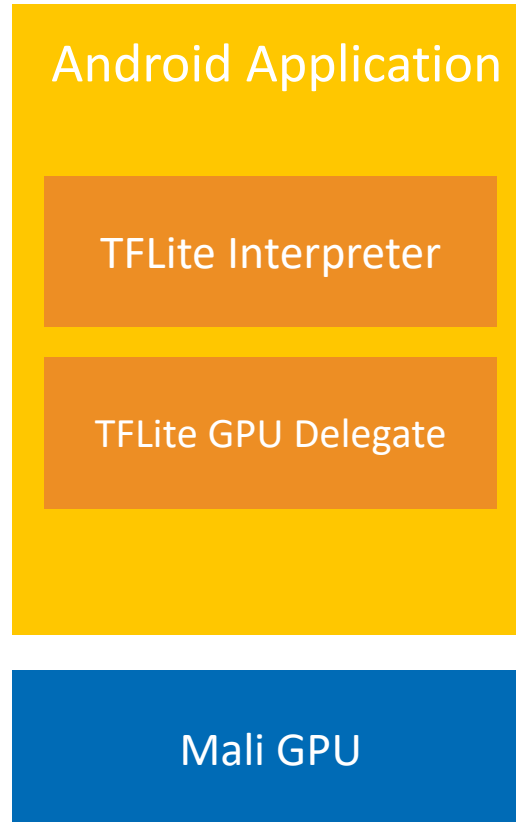


Pose Estimation



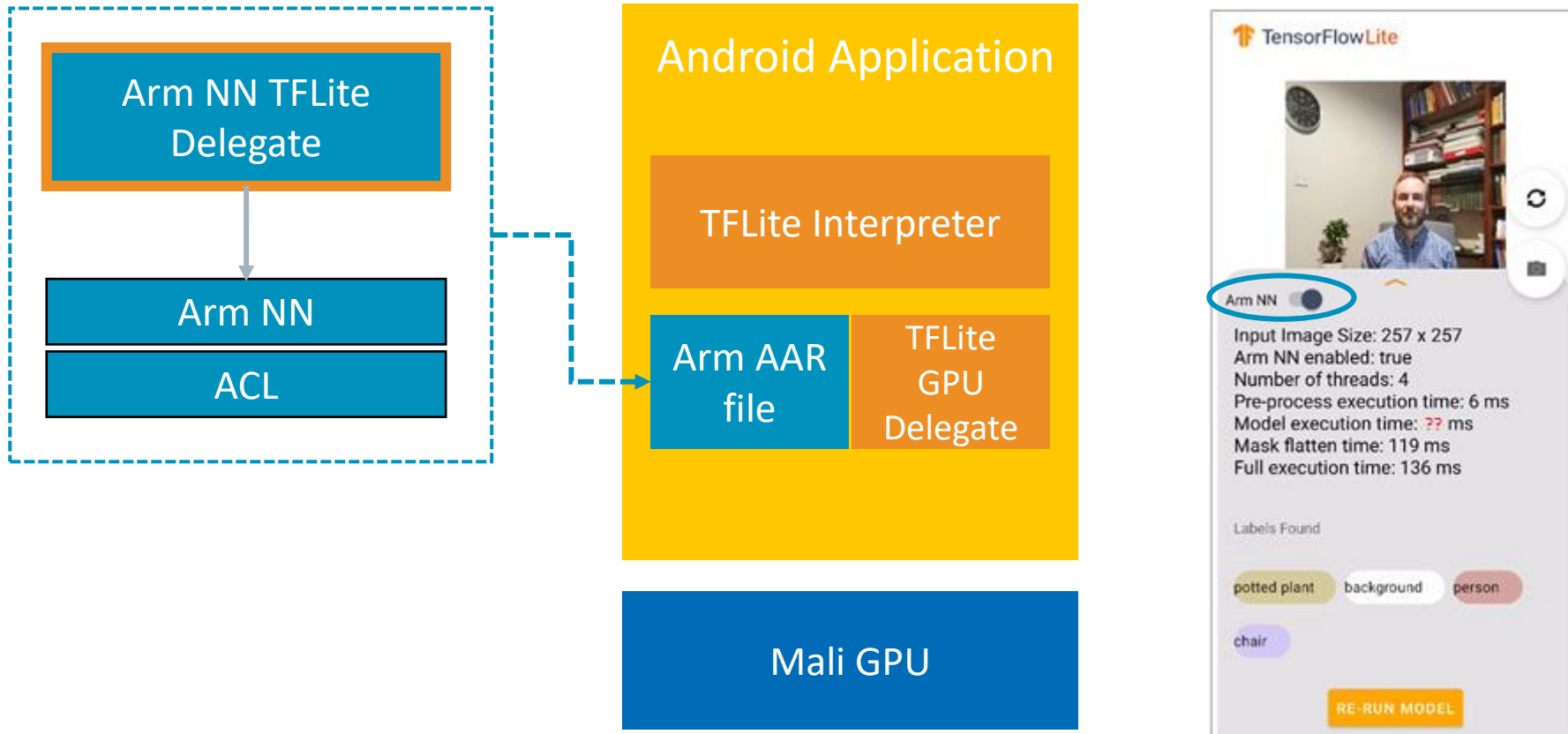
Android Demo – Image Segmentation

Using TensorFlow Lite GPU Delegate



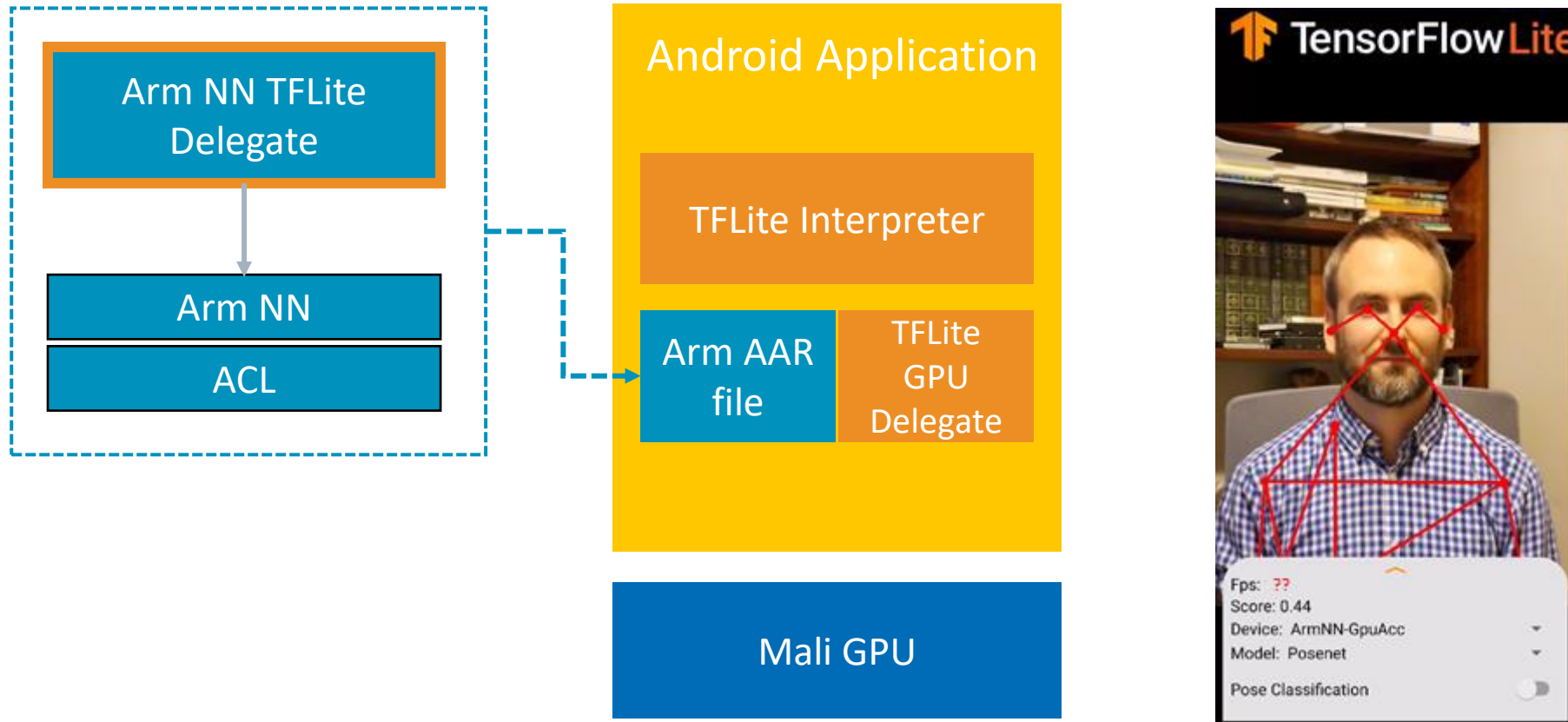
Android Demo – Image Segmentation

Using Arm NN and Arm Compute Library



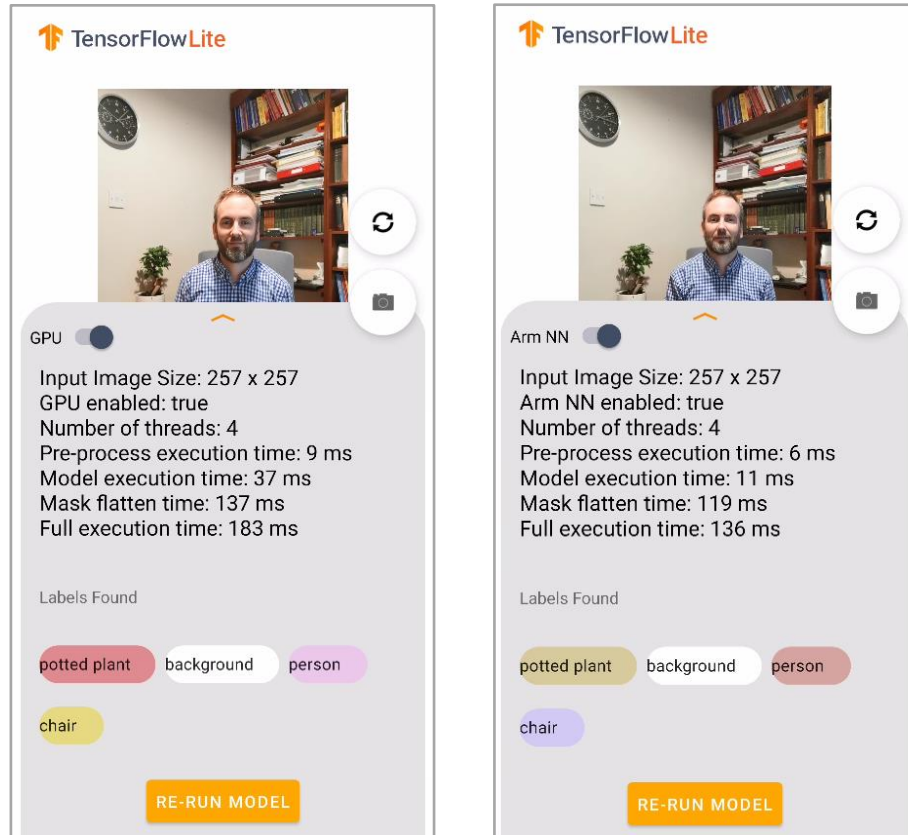
Android Demo – Pose Estimation

Using Arm NN and Arm Compute Library



Performance Summary

Image Segmentation



Accelerator	Approximate Execution time
TFLite GPU Delegate	37ms
Arm NN Delegate	11ms

Pose Estimation



Accelerator	Average FPS achieved
TFLite GPU Delegate	17
Arm NN Delegate	24

What's next??

- Continuous performance improvement on CPU & GPU
- Support for new Arm technologies
- Addition to Maven Central
- Sync Fence support
- NNAPI Support Library interface
- AHardwareBuf support
- NNAPI FL6 features

arm AI

AI Virtual Tech Talks Series

Thank You

Danke

Merci

谢谢

ありがとう

Gracias

Kiitos

감사합니다

धन्यवाद

شكراً

תודה



The Arm trademarks featured in this presentation are registered trademarks or trademarks of Arm Limited (or its subsidiaries) in the US and/or elsewhere. All rights reserved. All other marks featured may be trademarks of their respective owners.

www.arm.com/company/policies/trademarks

arm

Thank you!

Tweet us: [@ArmSoftwareDev](https://twitter.com/ArmSoftwareDev) -> #AIVTT

Check out our Arm Software Developers YouTube [channel](#)

Signup now for our next AI Virtual Tech Talk: developer.arm.com/techtalks