

TVM Meetup - Linaro

Jammy Zhou
November 16th, 2019



Bringing together the Arm ecosystem



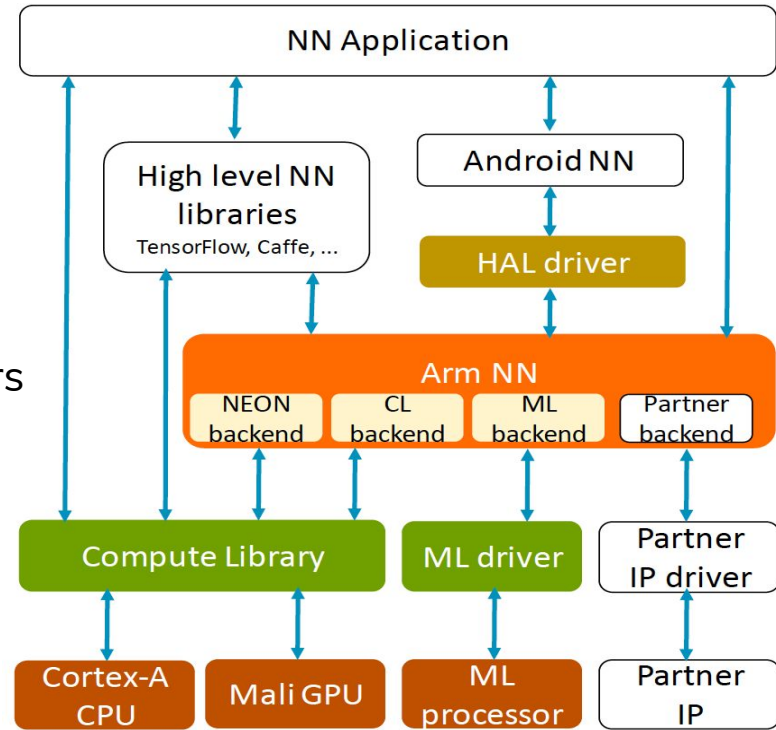
Linaro AI Initiative

Provide the best-in-class Deep Learning performance by leveraging Neural Network acceleration in IP and SoCs from the Arm ecosystem, through collaborative seamless integration with the ecosystem of AI/ML software frameworks and libraries

The logo for Arm, featuring the word "arm" in a blue, lowercase, sans-serif font.The logo for NXP, featuring the letters "NXP" in a stylized, blocky font with a yellow-to-green gradient.The logo for Qualcomm Innovation Center (QuIC), featuring a blue square with a white "Q" and the text "QUALCOMM INNOVATION CENTER, INC. QuIC" to its right.The logo for Socionext, featuring the word "socionext" in a lowercase, sans-serif font.The logo for STMicroelectronics, featuring a stylized "ST" in blue and the text "life.augmented" below it.The logo for Texas Instruments, featuring a red outline of the state of Texas with the text "TEXAS INSTRUMENTS" to its right.The logo for Amazon Web Services (AWS), featuring the word "aws" in a lowercase, sans-serif font with a yellow smile arrow below it.The logo for Amazon SageMaker Neo, featuring the text "Amazon SageMaker Neo" in a sans-serif font.The logo for TVM (TensorFlow Virtual Machine), featuring a blue square with a white "t" and the text "tvm" to its right.The logo for Google, featuring the word "Google" in its multi-colored, sans-serif font.The logo for TensorFlow Lite, featuring the TensorFlow logo (an orange arrow pointing up) and the text "TensorFlow Lite" in a sans-serif font.The logo for Microsoft, featuring the four-colored square logo and the word "Microsoft" in a sans-serif font.The logo for ONNX Runtime, featuring a stylized star logo and the text "ONNX RUNTIME" in a sans-serif font.

Arm NN open source project

- Linaro-hosted <https://www.mlplatform.org/>
- Git and review servers
- Forums and issue tracker
- Public mailing lists and IRC channel
- Internal [lira project](#) restricted to Linaro members
- Three sub-projects:
 - Arm Compute Library
 - Arm NN
 - Android NN Driver
- Arm Compute Library has been integrated by:
 - MATLAB Coder
 - ONNX Runtime



Arm platform support in TVM upstream

Green: Linaro 96Boards

IPs	Target	Hardware/Model	Options	Codegen
CPU	arm_cpu	pixel2 (snapdragon 835), mate10/mate10pro (kirin 970), p20/p20pro (kirin 970)	-target=arm64-linux-android -mattr=+neon	llvm
		firefly rk3399, rock960 , ultra96	-target=aarch64-linux-gnu -mattr=+neon	
		rasp3b (bcm2837)	-target=armv7l-linux-gnueabi -mattr=+neon	
		pynq	-target=armv7a-linux-eabi -mattr=+neon	
GPU	mali (midgard)	firefly rk3399, rock960 (mali t860)	N/A	opencl
	bifrost	hikey960 (mali g71)	N/A	
FPGA	vta	pynq, ultra96	N/A	sdaccel

Out-of-tree support or WIP: Hexagon DSP (via llvm), Ascend NPU, and more

Linaro for TVM

- Linaro AI/ML group can be a good fit for TVM collaborations on Arm based platforms to support more devices with various accelerator configurations (from microcontrollers to HPC) by working together with the members closely in an organized way
 - Arm - Cortex-A/Cortex-M/Neoverse CPU, Mali GPU, Ethos NPU
 - Qualcomm - Hexagon DSP, Adreno GPU
 - Hisilicon, Xilinx, NXP, TI, ST, Fujitsu, Riken, and etc
- Collaborations between Arm NN/ACL/CMSIS-NN and TVM
 - Integrate optimized ACL/CMSIS-NN kernels into TVM?
 - Implement Arm NN generic backend in TVM for more flexibility with the runtime plugins?
 - Integrate TVM codegen into Arm NN?
- CI and benchmark testing for TVM on member hardware platforms
 - Shall we maintain a list of Arm platforms supported by TVM?

More details from our members at the coming TVM Conference 2019, stay tuned...

Thank you

Join Linaro to accelerate deployment of your
Arm-based solutions through collaboration

contact@linaro.org

