RELEASE NOTES

Tensai Flow 1.0 Beta Patch Release

November 26, 2021

ETA COMPUTE
340 N Westlake Blvd. Suite #115, Westlake Village, California 91362

Table of Contents

1 Release Features

We call the software that runs on our device the “TENSAI platform Software”. In addition, we have software tools that run on a PC. “TENSAI Flow” is the combination of our “TENSAI Platform Software” and the Software Tools. This is the Beta patch version of TENSAI flow software. TENSAI flow consists of multiple software components and we will describe the features in terms of component capability in the next section.

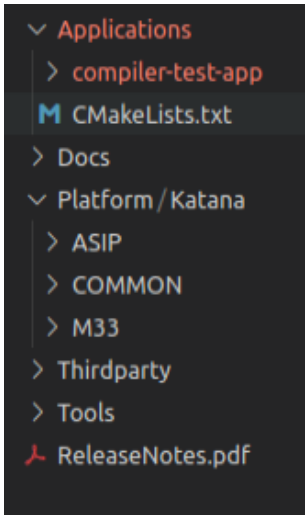
2 TENSAI Platform Software

Tensai platform Software consists of software components which abstracts the underlying low-level hardware interfaces from an application programmer. The middleware framework which we call “executor framework” frameworks allow scheduling NN kernels on different CPUs in a multicore environment, buffer management and synchronization in a multicore environment thus optimally using the computing resources and hardware acceleration.

3 Directory Structure

The top-level directory structure is defined below.

- Applications directory mainly contains all the sample applications
- Docs directory is placeholder for documentation
- Platform directory contains
    - M33 directory contains middleware and frameworks and M33 kernels
    - ASIP directory contains application for CAPE and LLE’s
    - COMMON directory contains common header files.
- Thirdparty directory contains the various other SW which is not owned by Eta Compute. These are SYNA SDK, FreeRTOS kernel, ARM CMSIS library and Kconfiglib, used extensively for our SW configuration.
- Tools directory contains the TENSAI compiler itself and other project generation and environment setup scripts.



4 TENSAI NN Zoo

TENSAI NN Zoo is a set of models trained in TensorFlow framework and quantized to int8 using TensorFlow Lite converter. With the current release, a simple model which can work with the camera as well as with auxiliary scripts to demonstrate the data conversion needed for testing the model.

- Fashion MNIST model : This model is trained on a dataset to recognize tiny images of apparel into 10 classes using Fashion MNIST dataset. The accuracy of this model is ~87.7% with 26.7 kilobytes of the weights. This model is not trained to work with camera images.
- MNIST model : This model is trained on a dataset to recognize tiny images of digits from 0-9, it will work with the test images provided along with the SDK. This model can be used to demonstrate the end to end pipeline working from camera to inference predictions. The accuracy of this model is ~95.7% on the test set with 17.9 kilobytes of the weights.
  Refer Docs/Running_MNIST_From_Camera.pdf on how to use this model with the camera.
- Synaptics memory test models : These are dummy models used for testing the memory usage constraints.

5 TENSAI Compiler

TENSAI Compiler is a tool to help the application developer port Intelligent Neural Network (NN) models to C code. The supported framework for NN models is Tensorflow Lite quantized with full-integer 8-bit quantization. It is distributed as a binary which can be run using command line without any software dependencies. The set of output files contains an inference function that calls the NN kernels and an empty template to integrate with the sensor data and post-processing the output. Compiler has capability to schedule the NN operations on two cores, LLE and CAPE2 or M33. It also allows scheduling operations on M33 or CAPE2 only if desired.

Below is a list of kernels that are supported by the tool:

- AveragePool 2D
- Concatenation
- Convolution 2D (including Pointwise Convolution)
- Depthwise Convolution 2D
- Elementwise Add
- Elementwise Multiply
- Maxpool 2D
- Fully connected (Dense)
- Reshape
- Softmax

For details refer
`Tools/Tensai_compiler/doc/TensaiKatanaCompiler_UserGuide.pdf`

6 Supported Devices

The release supports the Synaptics Katana KA10000 Rev A0 silicon.

7 Supported Boards

The release currently supports the following boards

- Tahiti EVK Board

8 Supported Hosts

The release can be built on following operating systems

- Ubuntu 18.04, 20.04 as primary OS
- Ubuntu 18.04, 20.04 on Virtual box VM with Windows Host

9 Supported Toolchain

The release supports and is tested using the following toolchain

- Build environment supported: Arm GNU tools gcc-arm-none-eabi-7-2017-q4-major
- Supported TensorFlow version: 2.4.1[1]

10 Changes

Command name changes from Tensai Flow Alpha2 to Tensai Flow 1.0 Beta.

- Move all include files into their corresponding “common_includes” and “includes” folder
- Build command for application changed from `make mcusdk image` to `make app`
- Tensai compiler command name changed from `tensai-nn-compile` to `tensaicc`

11 Enhancements

From Tensai Flow Alpha 2 to Tensai Flow 1.0 Beta

- Optimised NPU and CAPE kernels to reduce inference time
- Optimized executor to reduce inference time
- Change scripts to accept configs as run time parameters
- Added buffer dump and per layer timing analysis feature
- Added support to build LLE’s and CAPE using library
- Integrated Katana_SDK_Release_0.7.2
- Tool to get image inference via UART

From Tensai Flow Alpha 1 to Tensai Flow Alpha 2

- Optimised NPU kernels in terms of memory and inference time
- Optimized memory layout to fit larger models
- Load balancing between LLEA and LLEB to improve inference time

12 Fixes

From Tensai Flow Alpha 2 to Tensai Flow 1.0 Beta Patch

- Fix for “app demo crash after running some time”
- Softmax layer in one of the prediction heads for object detection type of models is now supported.
- Image pre processing bug fixes

[1] TensorFlow operators can change between versions, and other versions may cause undesired results.

13 Deprecated Features

NA

14 Known Issues and Problems

1. Popular architectures like MobileNet V1 and V2 use a GlobalAveragePooling2D layer before the final classification layer. TFLite will optimize this op as a 'Mean' instead of average pooling. However, Mean is an unsupported op in this release of Tensai Flow. A workaround is to explicitly use an AveragePooling2D layer.
2. Some deep models like MobileNet V2 may cause an unexpected overflow of the outputs, predicting a value of 127 when expected is -128. This typically happens when such deep models are initialized with default random weights in tensorflow as the scales and zero points become extremely tiny towards the prediction head. In general this problem won't occur for models that have already been trained.

15 Using the release

Please follow Quick Start Guide for using release

Docs/Quick_Start_Guide.pdf

16 Support

Contact Information

Address Eta Compute, Inc.
340 N Westlake Blvd. Suite 115
Westlake Village, California 91362

Website: https://etacompute.com
Support: support@etacompute.com

17 Legal Information and Disclaimers

ETA COMPUTE INTENDS FOR THE CONTENT CONTAINED IN THE DOCUMENT TO BE ACCURATE AND RELIABLE. THIS CONTENT MAY, HOWEVER, CONTAIN TECHNICAL INACCURACIES, TYPOGRAPHICAL ERRORS OR OTHER MISTAKES. ETA COMPUTE MAY MAKE CORRECTIONS OR OTHER CHANGES TO THIS CONTENT AT ANY TIME. ETA COMPUTE AND ITS SUPPLIERS RESERVE THE RIGHT TO MAKE CORRECTIONS, MODIFICATIONS, ENHANCEMENTS, IMPROVEMENTS AND OTHER CHANGES TO ITS PRODUCTS, PROGRAMS AND SERVICES AT ANY TIME OR TO DISCONTINUE ANY PRODUCTS, PROGRAMS, OR SERVICES WITHOUT NOTICE.

THE CONTENT IN THIS DOCUMENT IS PROVIDED "AS IS". ETA COMPUTE AND ITS RESPECTIVE SUPPLIERS MAKE NO REPRESENTATIONS ABOUT THE SUITABILITY OF THIS CONTENT FOR ANY PURPOSE AND DISCLAIM ALL WARRANTIES AND CONDITIONS WITH REGARD TO THIS CONTENT, INCLUDING BUT NOT LIMITED TO, ALL IMPLIED WARRANTIES AND CONDITIONS OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT OF ANY THIRD-PARTY INTELLECTUAL PROPERTY RIGHT.

ETA COMPUTE DOES NOT WARRANT OR REPRESENT THAT ANY LICENSE, EITHER EXPRESS OR IMPLIED, IS GRANTED UNDER ANY PATENT RIGHT, COPYRIGHT, MASK WORK RIGHT, OR OTHER INTELLECTUAL PROPERTY RIGHT OF ETA COMPUTE COVERING OR RELATING TO THIS CONTENT OR ANY COMBINATION, MACHINE, OR PROCESS TO WHICH THIS CONTENT RELATE OR WITH WHICH THIS CONTENT MAY BE USED.

USE OF THE INFORMATION IN THIS DOCUMENT MAY REQUIRE A LICENSE FROM A THIRD PARTY UNDER THE PATENTS OR OTHER INTELLECTUAL PROPERTY OF THAT THIRD PARTY, OR A LICENSE FROM ETA COMPUTE UNDER THE PATENTS OR OTHER INTELLECTUAL PROPERTY OF ETA COMPUTE.

INFORMATION IN THIS DOCUMENT IS PROVIDED SOLELY TO ENABLE SYSTEM AND SOFTWARE IMPLEMENTERS TO USE ETA COMPUTE PRODUCTS. THERE ARE NO EXPRESS OR IMPLIED COPYRIGHT LICENSES GRANTED HEREUNDER TO DESIGN OR FABRICATE ANY INTEGRATED CIRCUITS OR INTEGRATED CIRCUITS BASED ON THE INFORMATION IN THIS DOCUMENT. ETA COMPUTE RESERVES THE RIGHT TO MAKE CHANGES WITHOUT FURTHER NOTICE TO ANY PRODUCTS HEREIN. ETA COMPUTE MAKES NO WARRANTY, REPRESENTATION OR GUARANTEE REGARDING THE SUITABILITY OF ITS PRODUCTS FOR ANY PARTICULAR PURPOSE, NOR DOES ETA COMPUTE ASSUME ANY LIABILITY ARISING OUT OF THE APPLICATION OR USE OF ANY PRODUCT OR CIRCUIT, AND SPECIFICALLY DISCLAIMS ANY AND ALL LIABILITY, INCLUDING WITHOUT LIMITATION CONSEQUENTIAL OR INCIDENTAL DAMAGES. “TYPICAL” PARAMETERS WHICH MAY BE PROVIDED IN ETA COMPUTE DATA SHEETS AND/OR SPECIFICATIONS CAN AND DO VARY IN DIFFERENT APPLICATIONS AND ACTUAL PERFORMANCE MAY VARY OVER TIME. ALL OPERATING PARAMETERS, INCLUDING “TYPICALS” MUST BE VALIDATED FOR EACH CUSTOMER APPLICATION BY CUSTOMER’S TECHNICAL EXPERTS. ETA COMPUTE DOES NOT CONVEY ANY LICENSE UNDER NEITHER ITS PATENT RIGHTS NOR THE RIGHTS OF OTHERS. ETA COMPUTE PRODUCTS ARE NOT DESIGNED, INTENDED, OR AUTHORIZED FOR USE AS COMPONENTS IN SYSTEMS INTENDED FOR SURGICAL IMPLANT INTO THE BODY, OR OTHER APPLICATIONS INTENDED TO SUPPORT OR SUSTAIN LIFE, OR FOR ANY OTHER APPLICATION IN WHICH THE FAILURE OF THE ETA COMPUTE PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR. SHOULD BUYER PURCHASE OR USE ETA COMPUTE PRODUCTS FOR ANY SUCH UNINTENDED OR UNAUTHORIZED APPLICATION, BUYER SHALL INDEMNIFY AND HOLD ETA COMPUTE AND ITS OFFICERS, EMPLOYEES, SUBSIDIARIES, AFFILIATES, AND DISTRIBUTORS HARMLESS AGAINST ALL CLAIMS, COSTS, DAMAGES, AND EXPENSES, AND REASONABLE ATTORNEY FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PERSONAL INJURY OR DEATH ASSOCIATED WITH SUCH UNINTENDED OR UNAUTHORIZED USE, EVEN IF SUCH CLAIM ALLEGES THAT ETA COMPUTE WAS NEGLIGENT REGARDING THE DESIGN OR MANUFACTURE OF THE PART.