Technical Article

How to Simplify Your Embedded Edge AI Application Development

Manisha Agrawal

Designing a power-efficient edge artificial intelligence (AI) system while achieving a faster time to market can become tedious in the absence of the right tools and software from an embedded processor vendor. Challenges include selecting the right deep learning model, training and optimizing the model for performance and accuracy goals, and learning proprietary tools for model deployment on the embedded edge processor.

Starting from model selection to deployment on a processor, TI offers free tools, software and services designed to help you in each step of the deep neural network (DNN) development workflow, as shown in Figure 1. Let's walk through selecting a model, training your model anywhere, and deploying it seamlessly onto a TI processor without any hand-tooling, or manual programming, for hardware-accelerated inference.

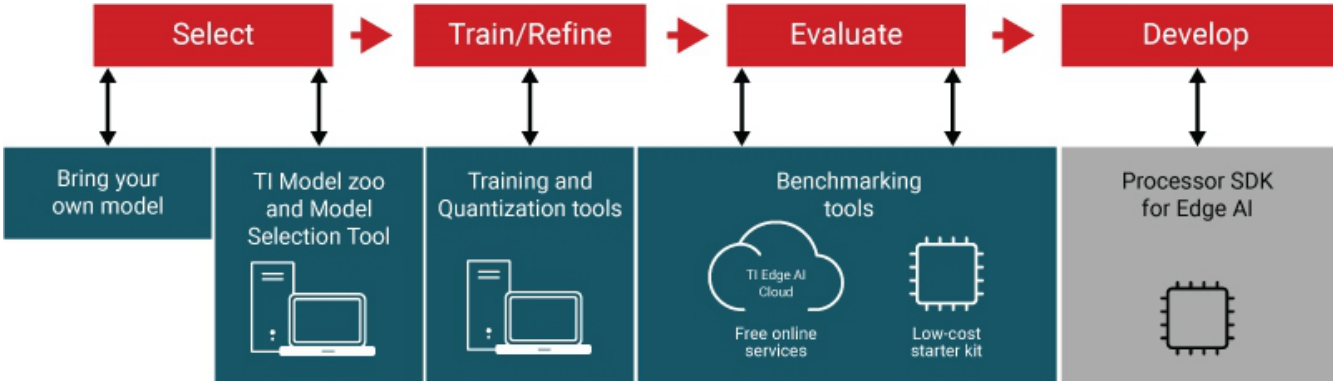

Figure 1. Edge AI application development workflow

Step No. 1: Selecting Your Model

The first task in edge AI system development is selecting the right DNN model while considering the performance, accuracy and power goals of your system. Tools such as the TI edge AI model zoo on GitHub and can help you accelerate this process.

The model zoo is a large collection of popular open-source deep learning models from TensorFlow, PyTorch and MXNet frameworks. The models are pre-trained on a public data set and optimized to run efficiently on TI processors for edge AI. TI regularly updates the model zoo with the latest models from the open-source community as well as TI-designed models to give you the widest choice of performance- and accuracy-optimized models.

With hundreds of models in the model zoo, the TI model selection tool, as shown in Figure 2, can help you quickly compare and locate the right model for your AI tasks by viewing and comparing performance statistics such as inference throughput, latency, accuracy and double-data-rate bandwidth without writing a single line of code.

Copyright © 2023 Texas Instruments Incorporated

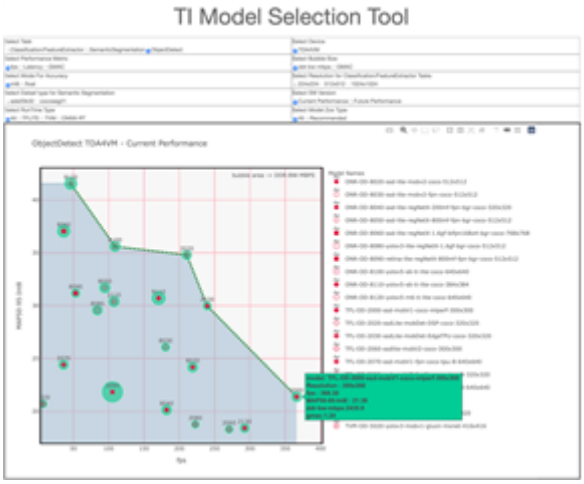

Figure 2. TI Model Selection Tool

Step No. 2: Training and Refining Your Model

After selecting your model, the next step is to train or refine the model for the best performance and accuracy on TI processors. Our software architecture and development environment lets you train your model anywhere.

When selecting a model from the TI model zoo, training scripts let you quickly transfer and train the models on the custom data set for your specific task without spending long cycles training from scratch or hand-tooling the model. For your own DNN models, training scripts, framework extensions and quantization-aware training tools help you optimize your model.

Step No. 3: Evaluating Model Performance

Evaluating model performance on actual hardware is required before developing an edge AI application.

TI's flexible software architecture and development environment let you train your model anywhere and compile and deploy it onto TI hardware with only few lines of code using your favorite industry-standard Python or C++ application programming interfaces (APIs) from TensorFlow Lite, ONNX RunTime or the TVM and SageMaker Neo with Neo AI DLR runtime engine. At the backend of these industry-standard runtime engines, our TI Deep Learning (TIDL) model compilation and runtime tools let you compile your model for TI hardware, deploy the compiled graphs or subgraphs onto the deep learning hardware accelerator, and get the best inference performance from the processor without any hand-tooling.

During the compilation step, the post-training quantization tool enables automatic conversion of floating-point models to fixed-point models. This tool enables layer-level mixed precision quantization (8 and 16 bit) through the configuration file, giving sufficient flexibility to tune the model compilation for the best performance and accuracy.

Operations across various popular models differ. The TI edge AI benchmark tool, also on GitHub, helps you match DNN model functionality seamlessly for models from the TI model zoo, and serves as a reference for custom models.

There are two ways to evaluate model performance on a TI processor: the TDA4VM starter kit evaluation module (EVM) or the TI Edge AI Cloud, a free online service that enables access remote access of the TDA4VM EVM for the evaluation of deep learning inference performance. Several example scripts for different tasks and runtime engine combinations help program, deploy and run accelerated inference on TI hardware in less than five minutes, while collecting benchmarks.

Get started with TI Edge AI Cloud

This free, open-source tool allows users to test their models using pretrained data models that are modifiable with user-uploaded data. Learn more about Edge AI Studio at dev.ti.com/edgeai.

Copyright © 2023 Texas Instruments Incorporated

Step No. 4: Developing Edge AI Applications

You can use open-source Linux® and industry-standard APIs to deploy your model onto TI hardware. Deploying your deep learning model onto a hardware accelerator is only a piece of the puzzle, however.

To help you quickly build an efficient edge AI application, TI has adopted the GStreamer framework. GStreamer plug-ins running on host Arm® cores let you automatically accelerate an end-to-end signal chain for computationally intensive tasks onto hardware accelerators and digital signal processing cores.

Figure 3 shows the software stack and components of the processor SDK with Linux for edge AI.

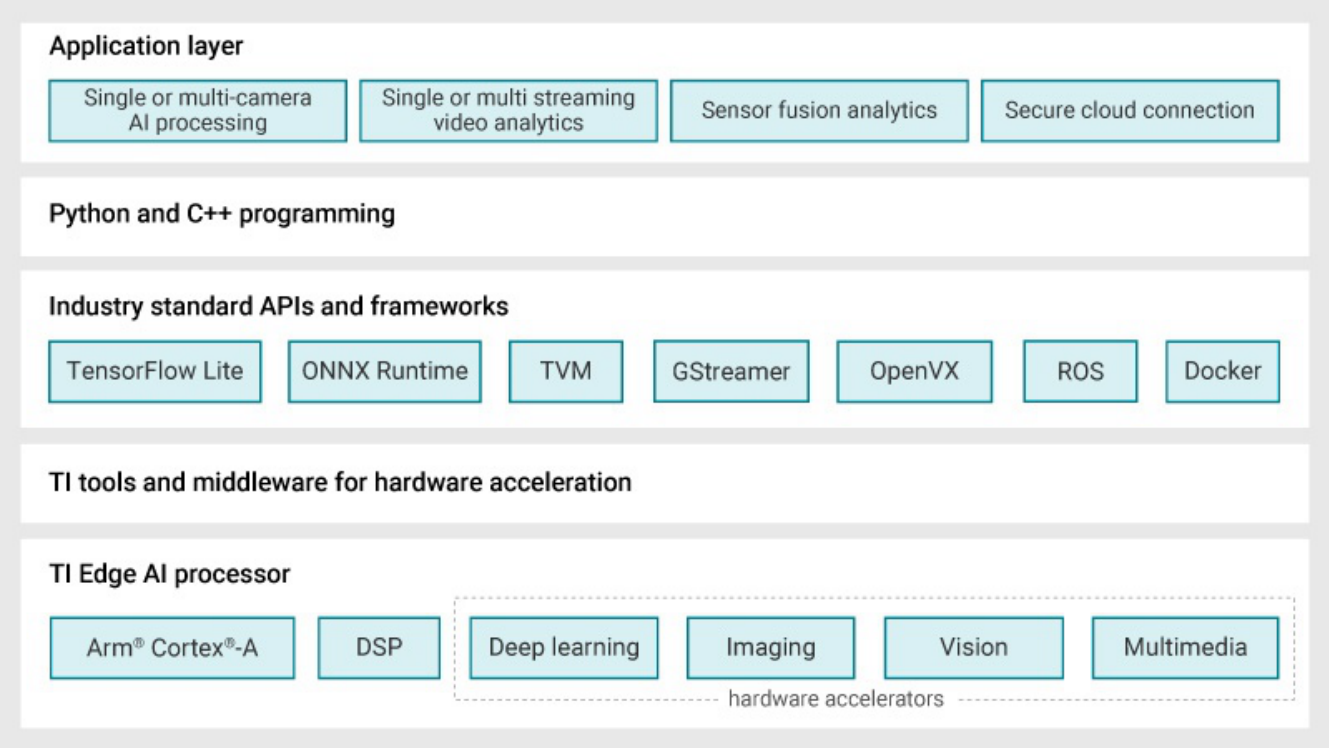

Figure 3. Processor SDK with Linux for Edge AI Components

Conclusion

If you are more of a novice and feel daunted by the tools that I've discussed in this article, rest assured that you do not need to be an AI expert to develop and deploy an AI model or build an AI application. The TI edge AI Academy helps you learn AI fundamentals and build an understanding of an AI system and software programming in a self-paced, classroom-style environment with quizzes. The labs present step-by-step code for building a "Hello, World" AI application, while an end-to-end advanced application with camera capture and display enable you to successfully develop an AI application at your own pace.

Additional Resources

- Learn more about enabling practical embedded intelligence.
- View documentation for edge AI tools

Copyright © 2023 Texas Instruments Incorporated

IMPORTANT NOTICE AND DISCLAIMER

TI PROVIDES TECHNICAL AND RELIABILITY DATA (INCLUDING DATA SHEETS), DESIGN RESOURCES (INCLUDING REFERENCE DESIGNS), APPLICATION OR OTHER DESIGN ADVICE, WEB TOOLS, SAFETY INFORMATION, AND OTHER RESOURCES “AS IS” AND WITH ALL FAULTS, AND DISCLAIMS ALL WARRANTIES, EXPRESS AND IMPLIED, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT OF THIRD PARTY INTELLECTUAL PROPERTY RIGHTS.

These resources are intended for skilled developers designing with TI products. You are solely responsible for (1) selecting the appropriate TI products for your application, (2) designing, validating and testing your application, and (3) ensuring your application meets applicable standards, and any other safety, security, regulatory or other requirements.

These resources are subject to change without notice. TI grants you permission to use these resources only for development of an application that uses the TI products described in the resource. Other reproduction and display of these resources is prohibited. No license is granted to any other TI intellectual property right or to any third party intellectual property right. TI disclaims responsibility for, and you will fully indemnify TI and its representatives against, any claims, damages, costs, losses, and liabilities arising out of your use of these resources.

TI’s products are provided subject to TI’s Terms of Sale or other applicable terms available either on ti.com or provided in conjunction with such TI products. TI’s provision of these resources does not expand or otherwise alter TI’s applicable warranties or warranty disclaimers for TI products.

TI objects to and rejects any additional or different terms you may have proposed.

Mailing Address: Texas Instruments, Post Office Box 655303, Dallas, Texas 75265
Copyright © 2023, Texas Instruments Incorporated